Journal Pre-proof

SARS-CoV-2 3CLpro whole human proteome cleavage prediction and enrichment/depletion analysis

Lucas Prescott

Please cite this article as: Lucas Prescott, SARS-CoV-2 3CLpro whole human proteome cleavage prediction and enrichment/depletion analysis, *Computational Biology and Chemistry,* (2021) doi:https://doi.org/10.1016/j.compbiolchem.2022.107671

SARS-CoV-2 3CLpro whole human proteome cleavage prediction and enrichment/depletion analysis

Lucas Prescott[1]*

[1] Independent Researcher, Boston, Massachusetts, United States of America

* Corresponding author
E-mail: lskywalker2015@gmail.com

Abstract

A novel coronavirus (SARS-CoV-2) has devastated the globe as a pandemic that has killed millions of people. Widespread vaccination is still uncertain, so many scientific efforts have been directed toward discovering antiviral treatments. Many drugs are being investigated to inhibit the coronavirus main protease, 3CLpro, from cleaving its viral polyprotein, but few publications have addressed this protease's interactions with the host proteome or their probable contribution to virulence. Too few host protein cleavages have been experimentally verified to fully understand 3CLpro's global effects on relevant cellular pathways and tissues. Here, I set out to determine this protease's targets and corresponding potential drug targets. Using a neural network trained on cleavages from 392 coronavirus proteomes with a Matthews correlation coefficient of 0.985, I predict that a large proportion of the human proteome is vulnerable to 3CLpro, with 4,898 out of approximately 20,000 human proteins containing at least one putative cleavage site. These cleavages are nonrandomly distributed and are enriched in the epithelium along the respiratory tract, brain, testis, plasma, and immune tissues and depleted in olfactory and gustatory receptors despite the prevalence of anosmia and ageusia in COVID-19 patients. Affected cellular pathways include cytoskeleton/motor/cell adhesion proteins, nuclear condensation and other epigenetics, host transcription and RNAi, ribosomal stoichiometry and nascent-chain detection and degradation, ubiquitination, pattern recognition receptors, coagulation, lipoproteins, redox, and apoptosis. This whole proteome cleavage prediction demonstrates the importance of 3CLpro in expected and nontrivial pathways affecting virulence, lead me to propose more than a dozen potential therapeutic targets against coronaviruses, and should therefore be applied to all viral proteases and subsequently experimentally verified.

Introduction

Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses with giant genomes (26-32 kb) that cause diseases in many mammals and birds. Since 2002, three human coronavirus outbreaks have occurred: severe acute respiratory syndrome (SARS) in 2002-2004, Middle East respiratory syndrome (MERS) from 2012 to present, and coronavirus disease 2019 (COVID-19) from 2019 to present. The virus that causes the latter disease, SARS-CoV-2, was first thought to directly infect the lower respiratory epithelium and cause pneumonia in susceptible individuals. The most common symptoms include fever, fatigue, nonproductive or productive cough, myalgia, anosmia, ageusia, and shortness of breath. More recently, however, correlations between atypical symptoms (chills, arthralgia, diarrhea, conjunctivitis, headache, dizziness, nausea, severe confusion, stroke, and seizure) and severity of subsequent respiratory symptoms and mortality have motivated researchers to investigate additional tissues that may be infected. One way to explain these symptoms and associated cellular pathways is to

review enrichment and depletion in virus-host interaction networks, particularly those including the coronavirus proteases.

Angiotensin-converting enzyme 2 (ACE2), the main receptor for SARS-CoV-1 and -2, has been shown to be less expressed in lung than in many other tissues. Respiratory coronaviruses likely first infect the nasal epithelium and tongue[1] and then work their way down to the lung and/or up through the cribriform plate to the olfactory bulb, through the rhinencephalon, and finally to the brainstem.[2-5] Additionally, based on ACE2 expression and *in vitro* and *in vivo* models, multiple parts of the gastrointestinal tract (mainly small and large intestine, duodenum, rectum, and esophagus; less appendix and stomach) and accessory organs (mainly gallbladder, pancreas, liver,[6, 7] salivary gland[8]; less tongue and spleen),[9] kidney,[10] male and female reproductive tissues,[11, 12] heart,[13] immune cells,[14, 15] and adipose tissue[16-18] may be infectible with corresponding symptoms and comorbidities.

Coronaviruses have two main open reading frames, orf1a and orf1b, separated by a ribosomal frameshift and resulting in two large polyproteins, pp1a and pp1ab, containing proteins including two cysteine proteases,[19] an RNA-dependent RNA polymerase, and other nonstructural proteins (nsp1-16). The main function of these proteases is to cleave the polyproteins into their individual proteins to form the transcription/replication complex, making them excellent targets for antiviral drug development.[20-23] The papain-like protease (PLpro) and 3 chymotrypsin-like protease (3CLpro) only have 3 and 11 cleavage sites, respectively, in the polyproteins, but it is reasonable to assume that both proteases may cleave host cell proteins to modulate the innate immune response and enhance virulence as in picornaviruses and retroviruses, such as human immunodeficiency virus (HIV).

PLpro is a highly conserved protein domain that has been shown to determine virulence of coronaviruses[24] and possess deubiquinating and deISGylating activity including cleaving ISG15 induced by interferon via the JAK-STAT pathway from ubiquitin-conjugating enzymes and potentially from downstream effectors.[25-29] PLpro deubiquination also prevents activating phosphorylation of IRF3 and subsequent type I interferon production,[30, 31] however the ubiquitinated leucine in human IRF3 is replaced by a serine in bats likely including *Rhinolophus affinus* (intermediate horseshoe bat), the probable species of origin of SARS-CoV-2.[32, 33]

3CLpro is also highly conserved among coronaviruses; SARS-CoV-2 3CLpro is 96.08% and 50.65% identical, respectively, to the SARS- and MERS-CoV homologs, the former with only 12 out of 306 amino acids substituted with all 12 outside the catalytic dyad or surrounding pockets.[34-36] Even the most distant porcine deltacoronavirus HKU15 3CLpro shares only 34.97% identity yet is similarly conserved in the these important residues. This conservation indicates that all these proteases are capable of cleaving similar sequences no matter the protease genus of origin. In addition to the 11 sites in the polyproteins, these proteases are known to cleave host proteins including STAT2,[37] NEMO,[38] NLRP12, and TAB1[39] to modulate interferon signaling. Similar proteases have been studied in the other members of *Nidovirales*[40] and the related *Picornavirales*,[41-45] with foot-and-mouth disease virus (FMDV) 3Cpro cleaving histone H3,[46, 47] poliovirus 3Cpro cleaving TFIID and TFIIIC,[48-52] and polio- and rhinovirus but not cardiovirus 3Cpro cleaving MAP4.[53, 54] These results, however, have not been reproduced for SARS-CoV-2 yet, and STAT2, NEMO, NLRP12, TAB1, H3, TFIIIC, TFIID, and MAP4 are only a few of many cleaved proteins.

The high number of 3CLpro cleavages in coronavirus polyproteins has, however, allowed for sequence logos and resulting sequence rules and training of decision trees and neural networks (NN) for additional cleavage site prediction.[55-60] Notably, Kiemer et al.'s NN[59] based on Blom et al.'s equivalent picornaviral NN[60] was trained on 7 arbitrary coronavirus genomes, totaling 77 cleavages, and had a Matthews correlation coefficient (MCC) of 0.84, much higher than the traditional consensus pattern's 0.37 for the same training set size. They predicted cleavage sites in select host proteins, namely the transcription factors CREB-RP, OCT-1, and multiple subunits of TFIID, the innate immune

2

modulators IFI6 and IRAK-1, the epithelial ion channels CFTR and SCNN1D, the tumor suppressors p53BP1/2 (although not p53 itself), RNA polymerase subunits RPA1 and RPC1, eIF4G1, the cytoskeletal proteins MAP4 and MAPRE1/3, and many members of the ubiquitin pathway (USP1/4/5/9X/9Y/13/26 and SOCS6).

Additionally, Yang's decision trees[58] were trained on 4 amino acid sliding windows and substitution matrix similarity score-based embeddings, achieved MCCs up to 0.95, but were limited to only 18 coronavirus polyproteins. The embedding-derived non-orthogonality somewhat stabilized the prediction to small changes in sequence assuming the substitution matrix reflects how the cleavages evolve. Decision trees have the benefit of being symbolic and explainable but often predict suboptimally when presented with interpolated or extrapolated inputs, making alternative machine learning techniques more attractive for predicting human protein cleavage prediction. For example, Narayanan et al.[61] and later Singh et al.[62] demonstrated that neural networks outperform decision trees for HIV and hepatitis C virus (HCV) protease cleavage prediction. Additional mixed methods such as Li et al.'s nonlinear dimensionality reduction and subsequent support vector machine (SVM) are able to retain some of the benefits of both linear and nonlinear classifiers.[63] Rognvaldsson et al.[64, 65] argue that nonlinear models including neural networks should not be used for cleavage prediction, however the HIV dataset from Cai et al.[66] that they used and their expanded dataset only included 299 and 746 samples, respectively. Additionally, physiochemical or structural encodings have outperformed one-hot encoding (also called orthogonal encoding) for their small HIV datasets[67] and have moreover eliminated differences between linear and nonlinear classifiers in an equivalent HCV dataset with 891 samples.[68] To my knowledge no one has expanded the 3CLpro cleavage dataset to the point where nonlinearity becomes significant, investigated the entire human proteome for 3CLpro cleavages sites with any method, or performed enrichment analysis and classification of these affected proteins.

Methodology

Dataset Preparation

A complete, manually reviewed human proteome containing 20,350 sequences (not including alternative isoforms) was retrieved from UniProt/Swiss-Prot (proteome:up000005640 AND reviewed:yes).[69]

Coronavirus polyprotein sequences were collected from GenBank.[70] Searching for "orf1ab," "pp1ab," and "1ab" within the family *Coronaviridae* returned 391 different, complete polyproteins, and an additional polyprotein sequence from the monotypic *Microhyla letovirus 1* was derived from accession number GECV01031551.[71] These polyproteins each contained 11 cleavages manually discovered using the Clustal Omega multiple sequence alignment server,[72-74] totaling 4,312 balanced cleavages (Figure 1). P1 glutamines and histidines were unambiguously conserved when aligned to known cleavages in SARS, SARS-CoV-2, MERS, IBV, etc., and all remaining glutamines and histidines were considered to be uncleaved. Although some of the ten amino acid sequences surrounding the cleavages were identical (805 different sites total), all 4,312 balanced positive cleavages were used for subsequent classifier training in addition to all other different, uncleaved sequences with P1 glutamines (18,477) and histidines (12,128), totaling 34,917 samples.
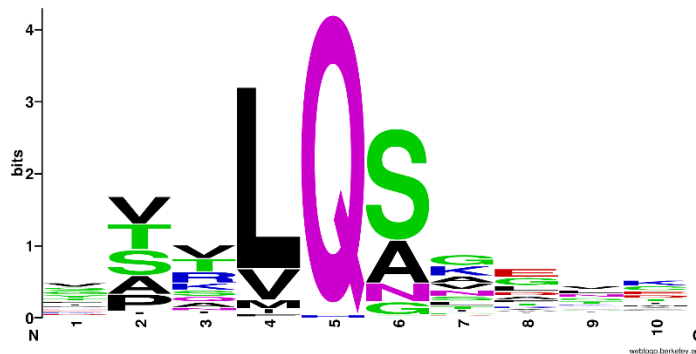
Figure 1: 3CLpro cleavage site sequence logo plotted by WebLogo v2.8.2.[75]

Dataset Characterization

Here I assumed that SARS-CoV-2 3CLpro is capable of cleaving all aligned cleavages between all genera of coronaviruses (*Alpha-*, *Beta-*, *Gamma-*, and *Deltacoronavirus* and the monotypic *Alphaletovirus*) because variation in cleavage sequences is greater within polyproteins than between them (Figures 2 and 3) no matter the existence of protease/cleavage cophylogeny (Figure 4).[76] Figures 1c and 1d demonstrate that the same 11 clusters appear when a lower-dimensionality physiochemical encoding (with dimensionality 40 containing normalized volumes, interface and octanol hydrophobicity scales, and isoelectric points) is used, however this dataset is large enough that one-hot encoding (200 dimensional binary input) outperforms it.
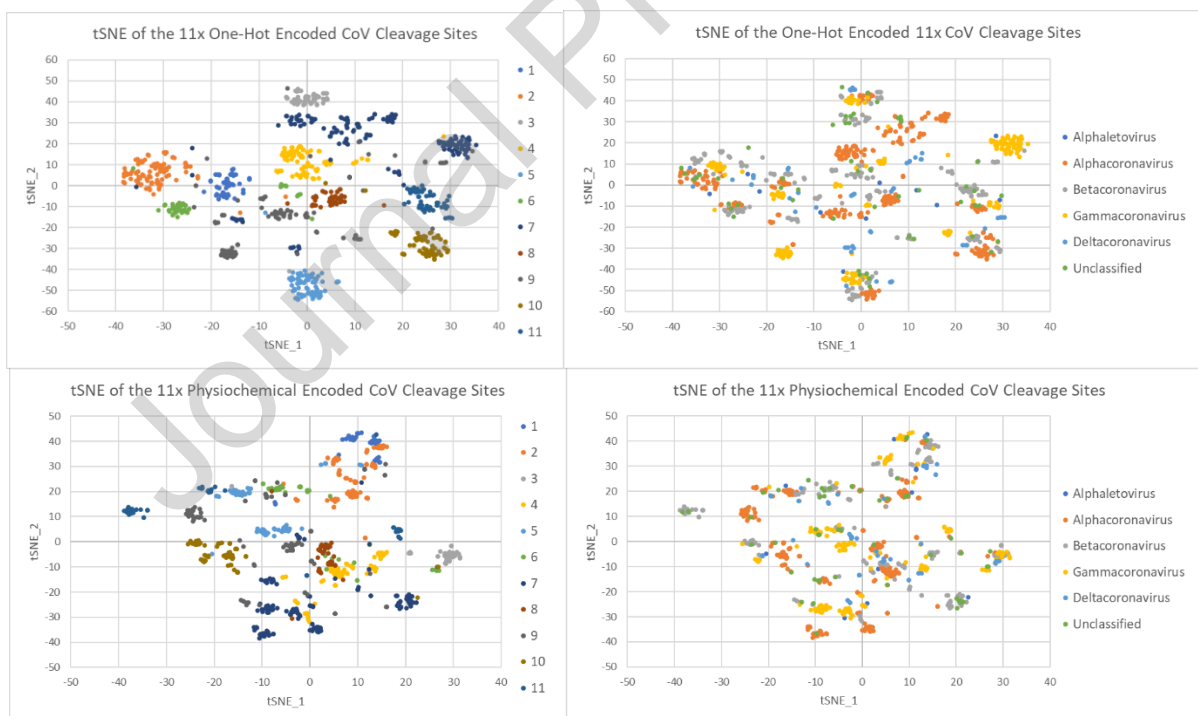
Figure 2: (A) One-hot encoded t-distributed stochastic neighbor embedding (t-SNE)[77] vs cleavage number and (B) vs genus demonstrating that variation within genomes is more important than variation between genomes. (C) Physiochemical encoded t-SNE vs cleavage number and (D) vs genus showing similar clustering although with worse separation.

4

Figure 3: (A) Information content vs residue position vs cleavage number and (B) vs genus similarly demonstrating that cleavage variation within single genomes is more important than variation between genomes.
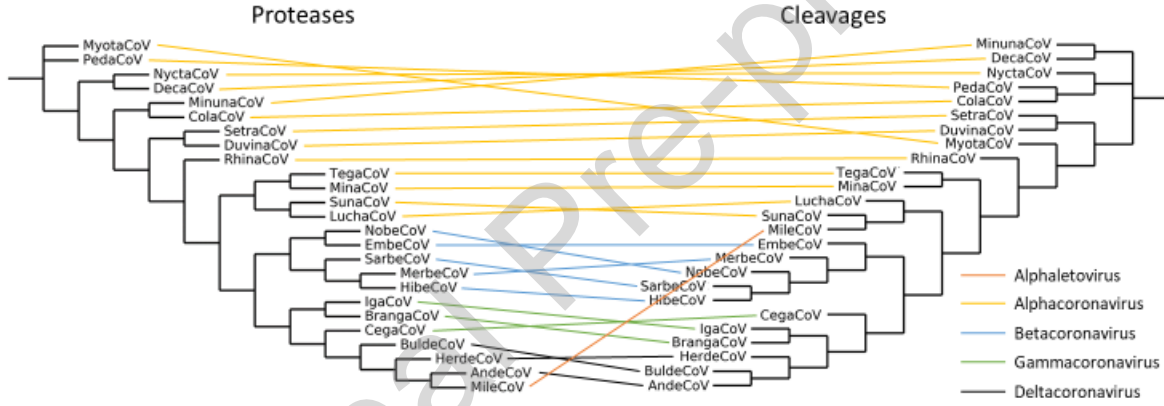
Figure 4: Unscaled subgenera-averaged tanglegram of 3CLpro and respective cleavages based on BLOSUM62 substitution matrix similarity scores with and without default affine gap penalties (opening 10 and extension 0.2).
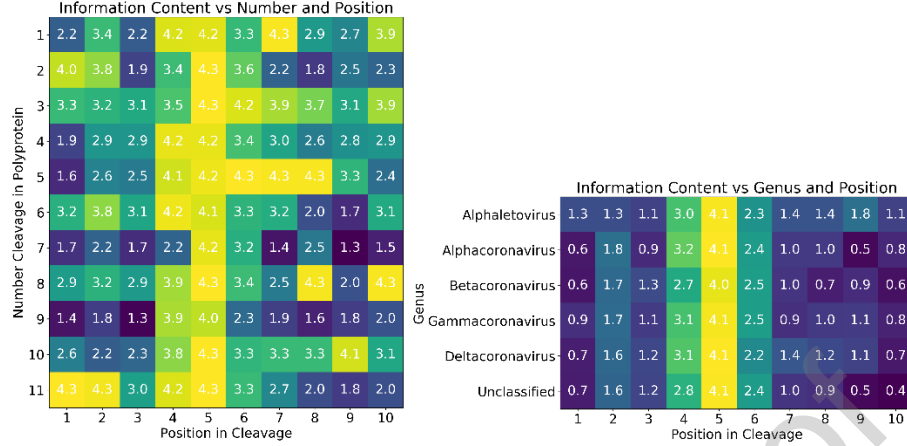
Model and Hyperparameter Optimization

The NetCorona 1.0 server as in Kiemer et al.'s work,[59] my reproductions of their sequence logo-derived rules and NN, and my improved sequence logo-based logistic regression and naïve Bayes classification and NNs were optimized and compared to decide which model to use for prediction of human cleavage sites.[78] Kiemer et al.'s seven genome sequence logo and multilayer perceptron used one-hot encoding for the 10 amino acid window surrounding each cleavage (linearizing 10 amino acids resulted in an input of 200 bits).[59] First, logistic regression was performed on the logit of the probability output of the sequence logo (as opposed to Chou et al.'s manual probability cutoff setting by maximizing an unbalanced measure of accuracy[79]) with a nonzero but optimally extremely small pseudocount and returned an MCC of 0.825 with 74.0% recall. Updating the sequence logo with all known cleavages (Figure 1) improved its MCC to 0.931 with 94.1% recall. A naïve Bayes classifier was additionally constructed from both the positive and negative sequence logos and slightly improved the MCC to 0.935 with 94.0% recall. Figure 5 demonstrates correlations (represented as the mutual information variant known as total entropy correlation coefficients or symmetric uncertainties) between positions that are not captured by simple sequence logos and classifiers assuming independence.[80, 81] NNs, however, allow inclusion of 2D and higher-order correlations not easily visualizable and therefore

5

often improve accuracy. Finally, in addition to information content, Figure 6 shows a charge-polarity-hydrophobicity scale with no obvious trend, reaffirming why one-hot encoding performs can achieve a higher MCC than any physiochemical, lower-dimensional encoding for NNs when the training set is large enough.
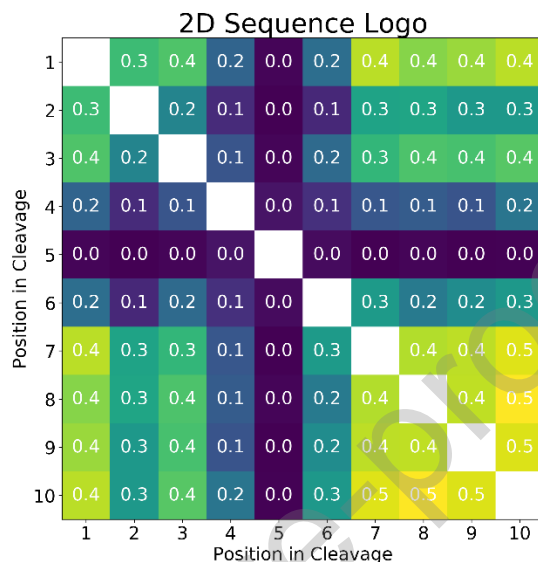
Figure 5: Entropy correlation coefficients (also known as symmetric uncertainties) between positions within the improved sequence logo.
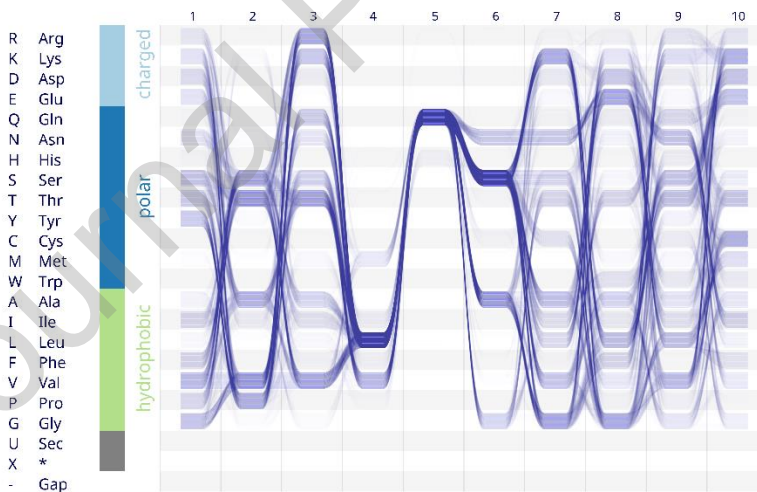
Figure 6: Sequence bundle with charge-polarity-hydrophobicity encoding.[82]

As for my improvements to the NN, note that Kiemer et al.'s MCC of 0.840 is an average from triple cross-validation (CV).[59] Because the known cleavage dataset is small, no data went unused; the three NN output scores were averaged and similarly considered cleavages when greater than 0.5. Retraining the same NN structures (each with one hidden layer with 2 neurons) on the larger dataset resulted in three-average CV MCC of 0.968, a significant improvement even though the datasets are less balanced. This MCC was maintained after adding all other histidines (which precede 20/805 different cleavages) as negatives. Interestingly, two infectious bronchitis viruses (*Igacovirus, Gammacoronavirus*) and one wigeon coronavirus HKU20 (*Andecovirus, Deltacoronavirus*) contained cleavages following leucine, methionine, and arginine (VSKLL^AGFKK in APY26744.1, LVDYM^AGFKK and DAALR^NNELM in

6

ADV71773.1, and AIRCR^NNELM in YP_005352870.1). To my knowledge, synthetic tetra/octapeptides have been cleaved following histidine, phenylalanine, tryptophan, methionine, and possibly proline residues,[56, 83] but only one natural histidine substitution has been documented in HCoV-HKU1[84] and likely does not affect function.[85-87]

To optimize hyperparameters, the whole dataset was repeatedly split into 80% training/20% testing sets with further splitting of the 80% training set for cross-validation. The optimal settings, no oversampling (within training folds[88]), limited-memory Broyden-Fletcher-Goldfarb-Shanno (lbfgs) solver, rectifier (ReLU) activation, 0.00001 regularization, and 1 hidden layer with 10 neurons, had an average 20% test set MCC of 0.976 when split and trained many times. Train/test sets repeatedly split with different ratios in Figure 7 demonstrate that the entire dataset is not required for adequate performance for all three classification methods, although my final method used all the data to maximize accuracy. Note that any errors in these predictions are amplified when applied to the whole human proteome as below but that enrichment/depletion statistics proved robust against this variability. Similarly careful optimization and bias and variance characterization should again be performed if this type of analysis is to be repeated on other protease datasets. Also note that Figure 7 displays a curve for a physiochemical encoding (also used in Figures 1c and 1d) underperforming when compared to one-hot encoding even at relatively small training sizes. Of the four physiochemical scales used, octanol hydrophobicity alone reached an MCC of 0.959, and, in the order of importance, addition of volume, interface hydrophobicity, and isoelectric point features increased the maximum MCC to 0.977.
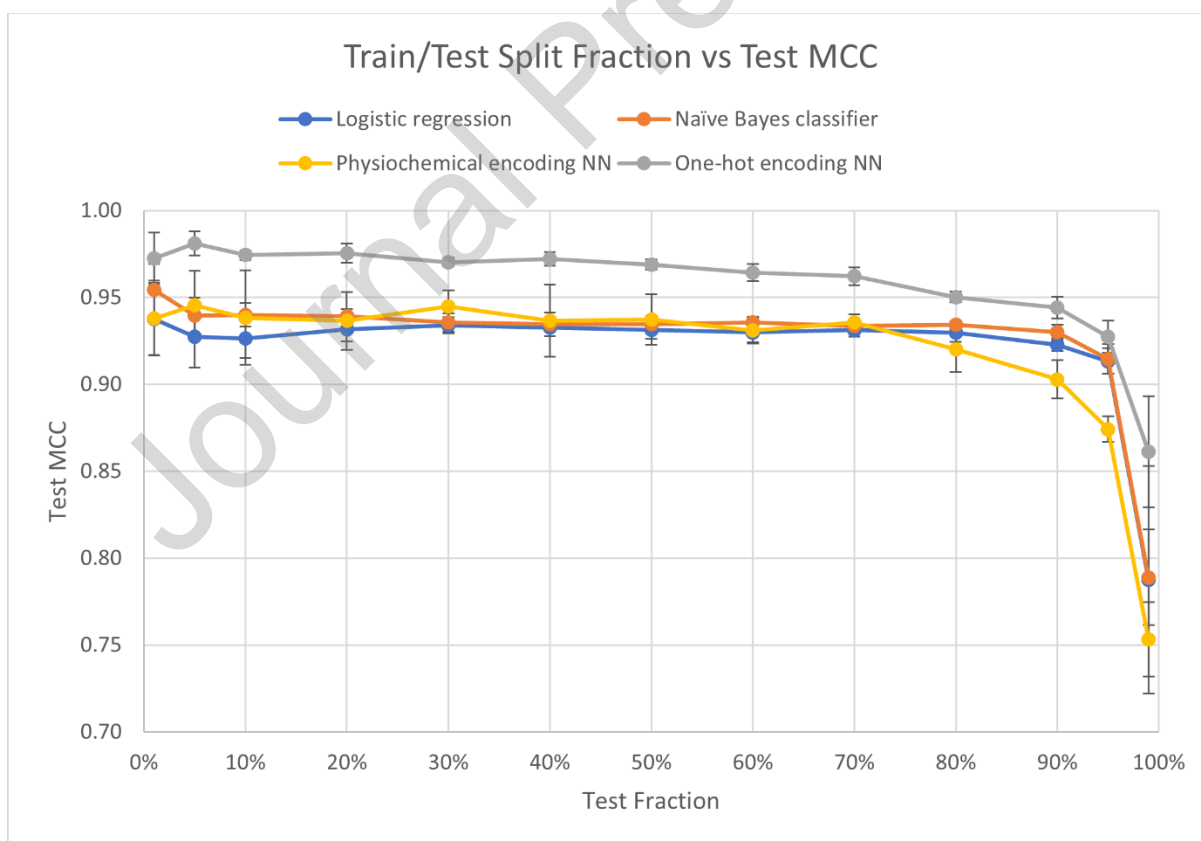

Figure 7: Random train/test split fraction vs MCC demonstrating that performance quickly approaches a limit for all classifiers.

7

Given that protease cleavage datasets are relatively small and training individual models is computationally inexpensive, combining multiple models into ensembles is recommended to reduce variability and at least slightly improve accuracy. The cross-validation described above is itself an ensemble that improves accuracy by introducing diversity in resampling like bootstrap aggregating. In addition to resampling methods, averaging ensembles of networks trained on the same dataset but initialized differently were able to improve accuracies as recently discovered in benchmark datasets.[89] Without an obvious upper bound on ensemble complexity, the final model used for subsequent analyses was an average from 10 sets of 100-fold cross-validated networks. The extremely few sequences incorrectly labeled varied with retraining and were not overrepresented in any lineage; in essence there was no distinction between easy-to-learn and hard-to-learn samples. The average 20% test MCC of this size ensemble was 0.985, although the final ensemble used the entire dataset.

Model Robustness

Even with the extremely high accuracies of models trained on this large dataset, randomly train/test splitting does not account for any taxonomic biases. One can easily imagine that extending this training dataset to the entire order *Nidovirales* or even the class *Pisoniviricetes* may not improve SARS-CoV-2 protease prediction without some (co)phylogenetic weighting or complex resampling algorithms and experimental verification. A novel leave-one-(sub)genus-out resampling analysis (using the final NN architecture and one-hot encoding) summarized in Table 1 affirms that more divergent lineages are more difficult to accurately predict, but that leaving out whole *Sarbecovirus* and *Betacoronavirus* resulted in the MCCs 0.865 and 0.835, still rivaling accuracies in previous publications. Alternatively, initially training on only *Sarbecovirus* sequences and progressively expanding the training set phylogenetically to *Milecovirus* only reduced *Sarbecovirus*-specific MCCs from 0.996 to 0.989 while increasing all other subgenera-specific MCCs to similar values. This again affirms that the entire dataset should be used and that diversity between the 11 cleavages is more important than between lineage.

Table 1: Leave-one-(sub)genus-out resampling vs number of samples and MCCs showing that complete removal of distinct sequences of any lineage from the training set can reduce that lineage's test accuracy independent of the number of number of positives and negatives or their imbalance.

Genus	Subgenus	Positives	Negatives	% Positive	MCC
Alphaletovirus	All/MileCoV	11	236	4.5%	0.628
	All	1241	9207	11.9%	0.881
	ColaCoV	139	337	29.2%	0.986
	DecaCoV	287	1108	20.6%	0.990
	DuvinaCoV	246	905	21.4%	0.966
	LuchaCoV	68	374	15.4%	0.985
	MinaCoV	239	661	26.6%	0.973
	MinunaCoV	243	772	23.9%	0.985
AlphaCoV	MyotaCoV	179	453	28.3%	0.986
	NyctaCoV	267	871	23.5%	0.990
	PedaCoV	360	1147	23.9%	0.977
	RhinaCoV	216	487	30.7%	0.986
	SetraCoV	217	852	20.3%	0.964
	SunaCoV	114	583	16.4%	0.837
	TegaCoV	365	1716	17.5%	0.891
	All	1200	10433	10.3%	0.835
	EmbeCoV	422	2741	13.3%	0.925
BetaCoV	HibeCoV	42	378	10.0%	0.973
	MerbeCoV	330	3346	9.0%	0.829

	NobeCoV	142	1488	8.7%	0.938
	SarbeCoV	381	2600	12.8%	0.865
	All	1761	5724	23.5%	0.884
GammaCoV	BrangaCoV	163	321	33.7%	0.985
	CegaCoV	45	337	11.8%	0.943
	IgaCoV	1706	5157	24.9%	0.892
	All	111	2078	5.1%	0.799
DeltaCoV	AndeCoV	11	352	3.0%	0.664
	BuldeCoV	89	1403	6.0%	0.859
	HerdeCoV	11	369	2.9%	0.683

Cleavage Prediction

Some predicted cleavage sites were close enough to the N- and C-termini that the ten amino acid window input into the neural network was not filled. These sites with P1 glutamine residue less than four amino acids from the N-terminus or less than five amino acids from the C-terminus were omitted because although they may be within important localization sequences, their cleavage kinetics are likely significantly retarded by truncation.

Of the 20,350 manually reviewed human proteins, 4,898 were predicted to be cleaved at least once with a final average NN score greater than or equal to 0.5. To prove that the cleavages were nonrandomly distributed among human proteins, random sequences with weighted amino acid frequencies were checked for cleavages. Cleavages occurred at 1.28% of glutamines (4.77% of amino acids)[90] or every 1,640 amino acids in these random sequences. Most proteins are shorter than this and would, if randomly distributed, follow a Poisson distribution; this data's deviation from this distribution indicates that many cleavages are intentional.

Enrichment Analysis

Protein annotation, classification, and enrichment analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) 6.8.[91, 92] Tissue (UP_TISSUE and UNIGENE_EST_QUARTILE), InterPro, direct Gene Ontology (GO includes cellular compartment (CC), biological process (BP), and molecular function (MF)), Reactome pathways, sequence features, and keywords annotations were all explored, and only annotations with Benjamini-Hochberg-corrected p-values less than 0.05 were considered statistically significant. Both enriched and depleted (no cleavages) annotations are listed in Tables S2-S10, and my training data, prediction methods, and results can be found on GitHub (https://github.com/Luke8472NN/NetProtease).

Discussion

Enrichment and depletion analyses are often used to probe the importance of annotations in many disease states, yet quantification is not possible without experimentation. Table 2 summarizes cleavages within and hypotheses about noteworthy pathways, however many caveats exist. First, if a protein is central to a pathway, a single cleavage may be all that is required to generate equivalent downstream outcomes. Cleaved proproteins such as coagulation factors or complement proteins may even be activated by 3CLpro cleavage. Additional exhaustive analysis or inclusion of some measure of centrality is required to determine if any insignificantly enriched or depleted pathways are still affected at central nodes (as in false negatives). Second, protease-,[56] substrate sequence-,[76, 83, 93-95] substrate truncation-,[96] pH-, temperature-, inhibitor type and concentration-, and time after infection-dependent cleavage kinetics convert this classification problem into a regression problem. Cleavage rates among the 11 cleavages per pp1ab vary by at least 50-fold and are uncorrelated with the scores from the classifier described here, so these predictions assume that 3CLpro exists in high enough concentrations and for a long enough time that rate constants do not matter because cleavage reactions

are complete. Third, longer proteins are more likely to be randomly cleaved and may confound conclusions about annotations containing them. Cleavages in longer proteins (e.g. cytoskeletal or cell-cell adhesion components) are no less important than those in shorter sequences, and annotations containing proteins with multiple cleavages deviating from Poisson distributions are more likely due to highly conserved sequences than simply protein length. Lastly, convergent evolution within the host may also result in false positives and may be partially avoided by investigating correlations between domains, motifs, repeats, compositionally biased regions, or other sequence or structural similarities and other functional and ontological annotations. Ideally, a negative control proteome from an uninfectible species could prevent false positives, but coronaviruses are extremely zoonotic. Here, depletions in the human proteome are taken to be negative controls. Comparison with a bat proteome with deficiencies in many immune pathways, however, may show which human cleavages are unintentional or exerted little or no selective pressure before cross-species transmission.

Table 2: Summary of noteworthy cleavages and possible consequences.

Pathway	Cleaved proteins	Uncleaved proteins	Possible consequences
Neurodegeneration	APP, tau, eIF4G1, DNAJC13, huntingtin, FUS, ataxin-1		Neuropsychiatric symptoms
Olfaction/gustation	Olfactory receptors, some adenylate cyclases and PDEs	GNAL, GNAS, some adenylate cyclases and PDEs	Anosmia and ageusia via neural cAMP signaling and programmed cell death
Cytoskeleton	Microfilaments, intermediate filaments, microtubules, spectrin		Altered vesicle and virus trafficking, DMV formation
Motor proteins	Myosin, kinesin, dynein		Altered vesicle and virus trafficking
Cell adhesion	Integrins, immunoglobulins, cadherins, selectins		Disrupted barriers and inflammation
Ras superfamily	Rho, Rab, Ran, Arf	Ras	Altered vesicle and virus trafficking, disrupted barriers
Cilia	NPHP1/2/4/5/8, BBS1/4/8/9/12, ALMS1, CC2D2A, RP1, RPGRIP1, LCA5, PKD1/2, PKHD1, DNAAF2, IFT80, DNAH5/11, DNAI2, and RSPH6A		Ciliary dyskinesia, reduced mucociliary escalator effectiveness, and anosmia
Coatamers and adaptors	COPII (SEC24A/24B/31A/31B), retromer component VSP13B, synergin gamma, GGA1, LYST, and AP1/2/3/5	COPI, clathrin, caveolae, AP4	Altered vesicle and virus trafficking, DMV formation
Nucleus	NPC subunits, importins, exportins, lamins, chromatin remodeling proteins (HATs, HDACs, SMC proteins, separase, topoisomerase III alpha), some DNA methyltransferases and demethylases	CTCF, all other topoisomerases, some DNA methyltransferases and demethylases	Altered chromatin condensation
Ribosome	A few ribosomal proteins (RPL4/10 and RPS3A/19), nascent-chain detection and degradation (ZNF598, NEMF, and LTN1)	All other ribosomal proteins	Increased frameshifting
SRPs	SRP68/72kDa	SRP9/14/19/54kDa	Increased envelope protein membrane insertion and increased frameshifting
RNAi	DICER1, AGO2, PIWIL1/3		Disrupted antiviral RNAi
HERVs	Syncytin-1/2, PEG10, HERVK-5/6/7/8/9/10/18/19/21/24/25/113, HERVH-2q24.1/3, HHLA1, HERVFC1-1, HERVS71-1		Inflammation, inhibited RTs and/or integrases which may normally produce antiviral RNAi
Vault	MVP, TEP1, PARP4, TERT		Unknown
PRRs	TLR2/3/6/8, CLEC4G/H1/H2/4K/10A/12A/13C/13E/14A/16A, KLRC3/C4/F1, ACG1, collectin-12, neurocan core protein, FREM1, layilin, PKD1, E-selectin, and thrombomodulin, NLRC2/3/4/5 and NLRP1/2/3/6/12/14		Disrupted innate immune responses
Downstream of PRRs	RIP1/2, NF-κB p65 and p100 subunits, CFLAR, TRIF, IRF2/9, DAXX, PI3K/AKT pathway (PIK3CB/G/D, PIK3R2/5/6, PIK3C2A, n/i/eNOS, TSC1/2), mTOR pathway (mTOR itself, SREBP1, RICTOR, PRR5L, ULK1/2, RBCC1, lipin-3, GRB10, FOXO1/3), MAPK pathway (MAP4K2/4/5, MAP3K1/4/5/8/9/10/12/15/16/17/18/19/21, KSR2, MAPK7/13/15), TFs (c-Jun, ATF2/6, CREB1/3/BP, SP1,		Disrupted innate immune responses

10

	OCT1/2, HSF2/2BP/4/5/X1/X2, Pol I initiator UBTF, Pol II initiators TFIID, multiple subunits of SL1 (TBP, TBPL2, TAF4/4B/6), and MED1/12/12L/13/15/17/22/23/28, and Pol III initiators TFIIIC and SNAPC1/5)		
IFNs and receptors		All IFNs and their receptors	Uncleaved due to redundancy
Downstream of IFNs	STAT1/2/4/5B and ISGs (GBP1, OAS1, PML, mitoferrin-1/2, TRIM5)		Disrupted innate immune responses
Apoptosis	Pro-apoptotic (CASP2/4/5/14), anti-apoptotic (Bcl-2, BIRC1/2/3/6)		Unknown
Lipid transport and adipokines	APOA5/B/C4/L1/(a), CETP, MTTP, and LRP2/6, leptin, leptin receptor, IL-6	SR-B1	Correlations with dyslipidemia and cardiovascular disease
Ubiquitin-proteasome system	E3s (NEDD4/4L, SMURF1/2, WWP1, ITCH, HECW1), cullins, proteasomal subunits (PSMD1/3/4/5/8/11/14)	Ubiquitin	PLpro cleavages required for full understanding
Helicases	SKI and NEXT subunits (SKIV2L, SKIV2L2, TTC37, PAPD5, ZCCHC7, RBM7), exosome subunits (DIS3, DIS3L1), DHX36	DDX1/5	Enhanced viral RNA stability, yet both pro- and anti-viral helicases are cleaved
Coagulation	coagulation factors II, III, VIII, XII, XIII, plasmin(ogen), VWF, plasma kallikrein, kininogen-1, and fibronectin, PAI-2, megsin, A1AT, angiotensinogen, PZI, CBG, LEI, and HSP47		Unknown
Antiproteases	A2M and PZP near their bait regions	Other small antiproteases	Antiviral activity
Complement	C1/3/4/5		Unknown, but if complement is activated, NETosis and hypercoagulability
Redox-active centers	DUOX1/2, NOX5, XO	Antioxidants, selenocysteine synthesis	Inflammation, ROS

Tissues

As expected in this data, the most significant tissue enrichment of 3CLpro cleavages are in the epithelium, but central and peripheral nervous tissues are also affected due to their similar expression and enrichment of complex structural and cell junction proteins. It is noteworthy that major proteins associated with multiple neurodegenerative diseases (Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, and spinocerebellar ataxia type 1) are also predicted to be cleaved. Testis has somewhat similar expression to epithelium and brain, highly expresses ACE2, and is enriched in movement/motility- (subset of structural proteins) and meiosis-related (chromosome segregation) proteins, further increasing the likelihood that this tissue is infectible. Spleen, however, does not express much ACE2, and its enrichment is likely due to genes with immune function and mutagenesis sites. Proteins with greater tissue specificity (3rd quartile) show additional enrichments along the respiratory tract (tongue, pharynx, larynx, and trachea), in immune tissues (lymph node and thymus), and in other sensory tissues (eye and ear). Combining tissues, tobacco use disorder is the only significantly enriched disease, but acquired immunodeficiency syndrome (AIDS) and atherosclerosis were surprisingly depleted.

Cleavages are also surprisingly depleted in olfactory and gustatory pathways given the virus' ability to infect related cells and present as anosmia and ageusia. Olfactory receptors are transmembrane rhodopsin-like G protein-coupled receptors that, when bound to an odorant, stimulate production of cAMP via the G protein and adenylate cyclase. The G proteins GNAL and GNAS are not cleaved, and some but not all adenylate cyclases are cleaved, likely resulting in an increase in cAMP. cAMP is mainly used in these cells to open their respective ligand-gated ion channels and cause depolarization, but it is also known to inhibit inflammatory responses through PKA and EPAC. Multiple PDEs that degrade cAMP but not PDE4, the major PDE in inflammatory and immune cells, are cleaved. PDE4 inhibitors have been shown to reduce destructive respiratory syncytial virus-induces inflammation in lung,[96] but olfactory receptor neurons are quickly regenerated and sacrifice themselves when infected by influenza A virus.[98] The depletion in cleavages and resulting increase in cAMP in these

neurons is likely to inhibit their programmed cell death long enough for the virus to be transmitted through the glomeruli to mitral cells and the rest of the olfactory bulb. Tongue infection may have similar mechanisms, and herpes simplex virus has been shown to be transmitted to the brainstem through the facial and trigeminal nerves.[99]

Gene Ontology

Cleaved proteins are depleted in the extracellular space (except for structural collagen, laminin, and fibronectin mainly associated) and enriched in the cytoplasm and many of its components, indicating that the selective pressure for cleavage is weaker once cells are lysed and the protease is released. In the cytoplasm, the most obviously enriched sets are in the cytoskeleton, motor proteins, cell adhesion molecules, and relevant Ras GTPases, particularly in microtubule organizing centers (MTOCs) including centrosomes, an organelle central to pathways in the cell cycle including sister chromatid segregation. More specifically, cleavage of cilia-associated proteins may contribute to dyskinesia and reduced mucociliary escalator effectiveness associated with many respiratory viruses including HCoV-229E and SARS and their resulting bacterial pneumonias.[100, 101] Additionally, cilial dysfunction in olfactory cells in COVID-19 leads to anosmia, although the main reported mechanism is nsp13 (helicase/triphosphatase)-centrosome interaction.[102] Coiled coils account for many of these cleavages and are primarily expressed in corresponding cellular compartments in the epithelium, testis, and brain. Only the coronavirus nsp1, nsp13, and spike proteins have so far been shown to interact with the cytoskeleton,[103-105] although many other viruses including influenza A virus,[106] herpes simplex virus, rabies virus, vesicular stomatitis virus, and adeno-associated virus[107] also modulate the cytoskeleton.[108] In neurons, this allows for axonal and trans-synaptic transport of viruses which can often be inhibited but sometimes exaggerated by cytoskeletal drugs often used in oncology.[109-112]

Modulation of these structural and motor proteins is required for formation of the double-membrane vesicles surrounding replicase complexes[113, 114] and for egress. Similarly required for vesicular transport, the coatomer COPI, clathrin, and caveolae pathways are untouched by 3CLpro, but COPII components are likely cleaved due to their function in selecting cargo[115, 116] and contribution to membrane curvature preventing inward nucleocapsid engulfment.[117] Cleavage of many adaptor subunits often targeting degradation leaves only the poorly characterized AP4 or other unknown pathways to handle egress. Modulators of any of these vesicle trafficking pathways may be effective treatments for COVID-19.

The nucleus is enriched because its nuclear localization signals and scaffolding proteins are cleaved. Additionally, many nuclear pore complex proteins and importins/exportins associated with RNA transport are also cleaved. Lamins, which are cleaved by caspases during apoptosis to allow chromosome detachment and condensation, are also cleaved by 3CLpro. Chromatin-remodeling proteins including HATs often containing bromodomains, HDACs, SMC proteins also containing coiled coils, separase, and topoisomerase III alpha, but not CTCF nor any other topoisomerases are cleaved, complicating the effects on chromosome condensation and global gene expression. HDAC inhibitors have been shown to decrease or increase virulence depending on the virus,[118-122] and some but not all DNA methyltransferases and demethylases are cleaved, further complicating these effects. Viruses benefit from preventing programmed cell death and its corresponding chromosomal compaction in response to viral infection (pyknosis), but they also attempt to reduce host transcription by condensing chromosomes and reroute translation machinery toward their own open reading frames.[123, 124] Relatedly, 28S rRNA has been shown to be cleaved by murine coronavirus, and ribosomes with altered activity are likely directed from host to viral RNAs.[125] Ribosome cleavages are depleted here because they are required for viral translation, but the few ribosomal proteins that are cleaved tend to be more represented in monosomes, not polysomes,[126] indicating that ribosomes that initiate faster than they elongate are preferred because they likely frameshift more frequently, allowing for control of the stoichiometric ratio of pp1a and pp1ab.[127] If slower ribosomes are not directly more likely to

12

frameshift, they are still less likely to participate if frameshift-induced traffic jams, collision-stimulated translation abortion and splitting,[128] and subsequent 60S subunit obstruction sensing and nascent-chain ubiquitylation, which is especially noteworthy because multiple proteins involved this quality control are predicted to be cleaved.[129] Signal recognition particle (SRP) subunits 68/72kDa associated with the ribosome are also predicted to be cleaved, and the uncleaved SRP9/14kDa are known to encourage translation elongation arrest to allow translocation including transmembrane domain insertion (e.g. coronavirus envelope protein) and have been associated with frameshifts.[130-132] In fact, frameshifting is a highly enriched keyword in cleaved proteins mainly due to endogenous retroviral (ERV) elements, some of which can activate an antiviral response via pattern recognition receptors (PRRs).[133] Some also resemble reverse transcriptases and may, like the CRIPSR system in prokaryotes, be capable of copying coronavirus genomic RNA to produce an RNAi response via the similarly cleaved DICER and AGO.[134] If the latter is true, individuals with distinct ERV alleles and loci may differentially respond to SARS-CoV-2 infection and/or treatment, especially exogenous RNAi. Lastly, ribosomal proteins are also included in the nonsense-mediated decay (NMD) pathway, which is likely depleted in cleavages because NMD has been shown to be a host defense against coronavirus genomic and subgenomic RNAs' multiple ORFs and large 3' UTRs.[135] It was also shown that the nucleocapsid protein inhibits this degradation but often cannot protect newly synthesized RNAs early in infection. The selective pressure on 3CLpro may be reversed by this nucleocapsid inhibition and the preferential degradation of host mRNAs such that host resources can again be directed toward viral translation.

In addition to affecting large organelles, 3CLpro is predicted to cleave all known components of vault. Vault function has not been completely described, but it has known interactions with other viruses.[136-138] TERT, which is associated with vault TEP1 is also cleaved, but is more frequently reported to be activated by other viral infections and/or promote oncogenesis.[139]

Other common viral process proteins are enriched in the epithelium and adaptive immune cells, and those cleaved may affect the heat shock response and other small RNA processing. Lactoferrin, an antiviral protein that is upregulated in SARS infection,[140] is also cleaved, although one of its fragments, lactoferricin, has known antiviral activity.[141] Many PRRs, their downstream effectors, and related pathways (PI3K/AKT/mTOR, MAPK, and nitric oxide synthesis, where nitric oxide has conflicting effects on viral infection[142,143]) and transcription factors are cleaved, yet no interferons nor their receptors are cleaved likely due to their redundancy. Downstream of interferon, however, multiple STATs and ISGs are cleaved. Finally, complicating the effects of infection on apoptosis, cleavages in both pro-apoptotic caspases and in the anti-apoptotic Bcl-2 and inhibitors of apoptosis exist.

Other Pathways and Keywords

Lipoproteins are a depleted keyword, but multiple apolipoproteins, lipid transfer proteins, and their receptors are predicted to be cleaved and, other than the proapoptotic APOL1,[144] are associated with chylomicrons, VLDL, and LDL as opposed to HDL, indicating that lipoproteins may contribute to the correlations between COVID-19 symptom severity, dyslipidemia, and cardiovascular disease. It was recently discovered that SARS-CoV-2 spike protein binds cholesterol, allowing for association with and reduced serum concentration of HDL. These findings combined with the 3CLpro cleavages show an opportunity for HDL receptor inhibitor treatment, especially antagonists of the uncleaved scavenger receptor SR-B1.[145] Cleavage of the adipokines leptin, leptin receptor, and IL-6 provide a mechanism for COVID-19 comorbidity with obesity independent of lipoproteins and indicate another potential treatment: anti-leptin antibodies.[146, 147]

Ubiquitinating and deubiquitinating (DUBs) enzymes are most enriched in the epithelium and the nucleus, and cleavages exist in E3 ubiquitin ligases such as NEDD4, E3-supporting cullins, and DUBs such as proteasomal base and lid subunits, but not in ubiquitin itself. NEDD4 has been shown to enhance influenza infectivity by inhibiting IFITM3[148, 149] and Japanese encephalitis virus by inhibiting autophagy,[150] but its ubiquitination of many diverse human viruses promotes their egress. IFITMs

13

generally have antiviral activity (others include HIV-1,[151] dengue virus,[152] and filoviruses[153]), but its use as a treatment for COVID-19 should be carefully considered given its varying effects among other coronaviruses.[154, 155] SARS-CoV-2 has two probable NEDD4 binding sites: the proline-rich, N-terminal PPAY and LPSY[156] in the spike protein and nsp8, respectively. Although the former sequence is APNY and is likely not ubiquitinated in SARS-CoV, small molecule drugs targeting this interaction or related kinases may be useful treatments for COVID-19 as they have been for other RNA viruses.[157-159] Further research is required to compare these cleavages to the PLpro deubiquitinating activity and the specificity and function of distinct ubiquitin and other ubiquitin-like protein linkage sites.[160, 161]

Helicases make up approximately 1% of eukaryotic genes and are enriched in cleavages with many containing RNA-specific DEAD/DEAH boxes. Most viruses except for retroviruses have their own helicase (nsp13 in SARS-CoV-2) and multiple human RNA helicases have been shown to sense viral RNA or enhance viral replication.[162-164] SARS nsp13 and nsp14 have been shown to be enhanced by the uncleaved human DDX5 and DDX1, respectively,[165, 166] however subunits of the antiviral, RNA-degrading SKI and NEXT complexes and the catalytic subunit of the interacting exosome complex are cleaved. Additionally, DHX36 cleavage may be motivated by its importance in dsRNA sensing when complexed with DDX1 and DDX21, signaling through the similarly cleaved TRIF to type I interferons.[167] The remaining cleaved DEAD/DEAH-box helicases tend to interact with RIG-I-like receptor dsRNA sensing or are involved in ribosome biogenesis or translation initiation. Their varying proviral and antiviral activities make recommending possible therapeutic targets impossible without further characterization.[168]

The coagulation cascade contains many predicted cleavages (coagulation factors II, III, VIII (also an acute-phase protein secreted in response to infection), XII, XIII, plasmin(ogen), von Willebrand factor, plasma kallikrein, kininogen-1, and fibronectin), but it is not trivial to predict if these cleavages are similar enough to those in the normal pathway to be activating or inhibiting even though 3CLpro is structurally similar to factors IIa and Xa.[169] Additionally, multiple cleaved serpin suicide protease inhibitors (PAI-2, megsin, A1AT, and the less relevant angiotensinogen, PZI, CBG, LEI, and HSP47) are related to coagulation, hinting that 3CLpro may increase both thrombosis and fibrinolysis rates or result in dose-dependent effects.[170, 171] Angiotensinogen is, however, unrelated to coagulation and is cleaved far from its N-terminus, so its effects on the renin-angiotensin system remain unknown. The structurally similar A2M has a predicted cleavage outside its protease bait region, however, the addition of a missense mutation Q694S would allow cleavage at the same site as factor XIII without reducing protease trapping ability as much as large deletions.[172, 173] Additional support for this potential exogenous replacement includes presence of serine in the same position in PZP, which shares 71% identity with A2M and contains a neighboring GAG site resembling known PLpro cleavages in its primary bait region. Most other antiproteases, however, are too small to have many potential cleavage sites even though they are a very important response to respiratory virus infection. Serpin or alpha globulin replacement therapy or treatment with modified small, 3CLpro competitive inhibitors may be a useful treatment for COVID-19.[174]

In addition to coagulation factors, the complement system can induce expulsion of neutrophil extracellular traps (NETs) intended to bind and kill pathogens.[175] NETs, however, simultaneously trap platelets expressing tissue factor and contribute to hypercoagulability. The complement pathway is not obviously enriched, but many central proteins (C1/3/4/5) are or have subunits that are cleaved, indicating viral adaptation to the classical, alternative, and likely lectin pathways.[176-178] Neutrophilia and NET-associated host damage are known to occur in severe SARS-CoV-2 infection, so inhibitors of the pathway are currently in clinical trials: histone citrullination, neutrophil elastase, and gasdermin D inhibitors to prevent release and DNases to degrade chromatin after release.[179, 180] Complement inhibition would likely similarly reduce the risks of hypercoagulability and other immune-mediated inflammation associated with COVID-19, but effects may vary widely between sexes and ages.[181, 182]

14

Redox-active centers including proteins involved in selenocysteine synthesis are additionally depleted in cleavages likely because of their involvement in avoiding cell death and innate immune response. Respiratory viruses differentially modulate redox pathways, balancing lysis-enhanced virion proliferation and DUOX2-derived reactive oxygen species (ROS)-induced interferon response.[183] In addition to depleted antioxidant proteins, cleavage of DUOX1, NOX5, and XO, the former of which are upregulated in chronic obstructive pulmonary disease (COPD),[184] indicates that coronaviruses prefer to reduce oxidative stress in infected cells, contrary to most COVID-19 symptoms. Given the diversity of responses to respiratory virus infections, each proposed antioxidant should be thoroughly evaluated before being recommended as a treatment of COVID-19.

The impact of post-translational modifications on viral protease cleavage frequency remains uncharacterized. Glutamine and leucine, the two most important residues in the cleavage sequence logo, are rarely modified, but serine, the next most important residue, is the most frequently phosphorylated amino acid. Analysis of keywords showed enrichment of phosphoproteins and depletion of disulfide crosslinked, lipid-anchored, and other transmembrane proteins.

Lastly, the keywords polymorphism and alternate splicing were enriched, indicating that additional variability between cell lines and between individuals are likely. Once health systems are not so burdened by the quantity of cases and multiple treatments are developed, personalized interventions will likely differ significantly between individuals.

Conclusion

Many expected and novel protein annotations were discovered to be enriched and depleted in cleavages, indicating that 3CLpro is a much more important virulence factor than previously believed. 3CLpro cleavages are enriched in the epithelium (especially along the respiratory tract), brain, testis, plasma, and immune tissues and depleted in olfactory and gustatory receptors. Affected pathways with discussed connections to viral infections include cytoskeleton/motor/cell adhesion proteins, nuclear condensation and other epigenetics, host transcription and RNAi, coagulation, pattern recognition receptors, growth factor, lipoprotein, redox, ubiquitination, and apoptosis. These pathways point toward many potential therapeutic mechanisms to combat COVID-19: cytoskeletal drugs frequently used against cancer, modulators of ribosomal stoichiometry to enrich monosomes, upregulation of DICER1 and AGO1/2, exogenous lactoferrin and modified antiproteases including alpha globulins, upregulation of serpins potentially via dietary antioxidants, complement inhibition, reduction of LDL and inhibition of HDL receptor (e.g. by antagonizing SR-B1), anti-leptin antibodies, and downregulating NEDD4 or related kinases and upregulating IFITMs. Pathway components with more complex disruption that may also deliver therapeutic targets but require elucidating experimental results include PDEs, histone acetylation, nitric oxide, and vesicle coatomers. It is also worth further investigating how 3CLpro contributes if at all to the correlations between obesity and severity of infection or to viral induction of autoimmune and potentially oncological conditions.

Expansion of the training dataset to the whole order *Nidovirales* or class *Pisoniviricetes* may provide more diversity to improve classifying methods if additional protease/cleavage coevolution does not invalidate the assumption of cross-reactivity. Issues requiring *in vitro* and *in vivo* experimentation include characterization of cleavage kinetics, any functional differences between proteases, the molecular effects of post-translation modifications, and the individual and population effects of polymorphisms in cleavage sequences on susceptibility to or severity of infection. Even though many caveats exist without experimentation, similar prediction, enrichment/depletion analysis, and therapeutic target identification should be performed for every other viral protease.

Supplementary Information
S1 Table. List of abbreviations.
S2 Table. Significant UP_TISSUE enrichments and depletions.
S3 Table. Significant UNIGENE_EST_QUARTILE enrichments and depletions.
S4 Table. Significant InterPro enrichments and depletions.
S5 Table. Significant GO CC enrichments and depletions.
S6 Table. Significant GO BP enrichments and depletions.
S7 Table. Significant GO MF enrichments and depletions.
S8 Table. Significant Reactome pathway enrichments and depletions.
S9 Table. Significant sequence feature enrichments and depletions.
S10 Table. Significant keyword enrichments and depletions.

References
1. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, et al. Olfactory and gustatory dysfunction as a clinical presentation of mild to moderate forms of COVID-19: A multicenter European study. Eur Arch Otorhinolaryngol. 2020 Aug;277(8):2251–61. doi: 10.1007/s00405-020-05965-1 PMID: 32253535
2. Baig AM, Khaleeq A, Ali U, Syeda H. Evidence of the COVID-19 virus targeting the CNS: Tissue distribution, host-virus interaction, and proposed neurotropic mechanisms. ACS Chem Neurosci. 2020 Mar;11(7):995–8. doi: 10.1021/acschemneuro.0c00122 PMID: 32167747
3. Lau KK, Yu WC, Chu CM, Lau ST, Sheng B, Yuen KY. Possible central nervous system infection by SARS coronavirus. Emerg Infect Dis. 2004 Feb;10(2):342–4. doi: 10.3201/eid1002.030638 PMID: 15030709
4. Netland J, Meyerholz DK, Moore S, Cassell M, Perlman S. Severe acute respiratory syndrome coronavirus infection causes neuronal death in the absence of encephalitic in mice transgenic for human ACE2. J Virol. 2008 Aug;82(15):7264–75. doi: 10.1128/JVI.00737-08 PMID: 18495771
5. Li YC, Bai WZ, Hashikawa T. The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. J Med Virol. 2020 Mar;2020:1–4. doi: 10.1002/jmv.25728 PMID: 32104915
6. Zhang C, Shi L, Wang FS. Liver injury in COVID-19: management and challenges. Lancet Gastroenterol Hepatol. 2020 May;5(5):428–30. doi: 10.1016/S2468-1253(20)30057-1 PMID: 32145190
7. Chau TN, Lee KC, Yao H, Tsang TY, Chow TC, Yeung YC, et al. SARS-associated viral hepatitis caused by a novel coronavirus: Report of three cases. Hepatology. 2004 Feb;39(2):302–10. doi: 10.1002/hep.20111 PMID: 14767982
8. Liu L, Wei Q, Alvarez X, Wang H, Du Y, Zhu H, et al. Epithelial cells lining salivary gland ducts are early target cells of severe acute respiratory syndrome coronavirus infection in the upper respiratory tract of rhesus macaques. J Virol. 2011 Apr;85(8):4025–30. doi: 10.1128/JVI.02292-10 PMID: 21289121
9. Zhan J, Deng R, Tang J, Zhang B, Tang Y, Wang JK, et al. The spleen as a target in severe acute respiratory syndrome. FASEB J. 2006 Nov;20(13):2321–8. doi: 10.1096/fj.06-6324com PMID: 17077309
10. Naicker S, Yang CW, Hwang SJ, Liu BC, Chen JH, Jha V. The novel coronavirus 2019 epidemic and kidneys. Kidney Int. 2020 May;97:824–8. doi: 10.1016/j.kint.2020.03.001 PMID: 32204907
11. Fan C, Ding Y, Lu WL, Wang J. ACE2 expression in kidney and testis may cause kidney and testis infection in COVID-19 patients. Front Med (Lausanne). 2021 Jan;7:563893. doi: 10.3389/fmed.2020.563893

12. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet. 2020 Mar;395(10226):809–15. doi: 10.1016/S0140-6736(20)30360-3 PMID: 32151335

13. Zheng YY, Ma YT, Zhang JY, Xie X. COVID-19 and the cardiovascular system. Nat Rev Cardiol. 2020 Mar;17:259–60. doi: 10.1038/s41569-020-0360-5 PMID: 32139904

14. Dandekar AA, Perlman S. Immunopathogenesis of coronavirus infections: implications for SARS. Nat Rev Immunol. 2005 Dec;5(12):917–27. doi: 10.1038/nri1732 PMID: 16322745

15. Gu J, Gong E, Zhang B, Zheng J, Gao Z, Zhong Y, et al. Multiple organ infection and the pathogenesis of SARS. J Exp Med. 2005 Aug;202(3):415–24. doi: 10.1084/jem.20050828 PMID: 16043521

16. Jia X, Yin C, Lu S, Chen Y, Liu Q, Bai J, et al. Two things about COVID-19 might need attention. Preprints [Preprint]. 2020 Feb. doi: 10.20944/preprints202002.0315.v1

17. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, et al. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. Obesity. 2020 Apr;28(7):1196–9. doi: 10.1002/oby.22831 PMID: 32271993

18. Elliot JG, Donovan GM, Wang KCW, Green FHY, James AL, Noble PB, et al. Fatty airways: Implications for obstructive disease. Eur Respir J. 2019 Dec;54(6):1900857. doi: 10.1183/13993003.00857-2019 PMID: 31624112

19. Ziebuhr J Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. J Gen Virol. 2000 Apr;81(4):853–79. doi: 10.1099/0022-1317-81-4-853 PMID: 10725411

20. Baez-Santos YM, St. John SE, Mesecar AD. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. Antiviral Res. 2015 Mar;115:21–38. doi: 10.1016/j.antiviral.2014.12.015 PMID: 25554382

21. Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung SH. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. J Med Chem. 2016 Feb;59(14):6595–628. doi: 10.1021/acs.jmedchem.5b01461 PMID: 26878082

22. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLOS Biology. 2005 Nov;3(10):e428. doi: 10.1371/journal.pbio.0030324 PMID: 16128623

23. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. Science. 2003 Jun;300(5626):1763–7. doi: 10.1126/science.1085658 PMID: 12746549

24. Neimeyer D, Mosbauer K, Klein EM, Sieberg A, Mettelman RC, Mielech AM, et al. The papain-like protease determines a virulence trait that varies among members of the SARS-coronavirus species. PLOS Pathog. 2018 Sep;14(9):e1007296. doi: 10.1371/journal.ppat.1007296 PMID: 30248143

25. Barretto N, Jukneliene D, Ratia K, Chen Z, Mesecar AD, Baker SC. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquinating activity. J Virol. 2005 Dec;79(24):15189–98. doi: 10.1128/JVI.79.24.15189-15198.2005 PMID: 16306590

26. Yang X, Chen X, Bian G, Tu J, Xing Y, Wang Y, et al. Proteolytic processing, deubiquitinase and interferon antagonist activities of Middle East respiratory syndrome coronavirus papain-like protease. J Gen Virol. 2014 Mar;95(3):614–26. doi: 10.1099/vir.0.059014-0 PMID: 24362959

27. Bailey-Elkin BA, Knaap RCM, Johnson GG, Dalebout TJ, Ninaber DK, van Kasteren PB, et al. Crystal structure of the Middle East respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquinating activity to demonstrate its role in innate immune suppression. J Biol Chem. 2014 Dec;289:34667–82. doi: 10.1074/jbc.M114.609644 PMID: 25320088

28. Li SW, Lai CC, Ping JF, Tsai FJ, Wan L, Lin YJ, et al. Severe acute respiratory syndrome coronavirus papain-like protease suppressed alpha interferon-induced responses through downregulation of extracellular signal-regulated kinase 1-mediated signalling pathways. J Gen Virol. 2011 May;92(5):1127–40. doi: 10.1099/vir.0.028936-0 PMID: 21270289

29. Xing Y, Chen J, Tu J, Zhang B, Chen X, Shi H, et al. The papain-like protease of porcine epidemic diarrhea virus negatively regulates type I interferon pathway by acting as a viral deubiquitinase. J Gen Virol. 2013 Jul;94(7):1554–67. doi: 10.1099/vir.0.051169-0 PMID: 23596270

30. Matthews K, Schafer A, Pham A, Frieman M. The SARS coronavirus papain like protease can inhibit IRF3 at a post activation step that requires deubiquination activity. Virol J. 2014 Dec;11:209. doi: 10.1186/s12985-014-0209-9 PMID: 25481026

31. Devaraj SG, Wang N, Chen Z, Chen Z, Tseng M, Barretto N, et al. Regulation of IRF-3-dependent innate immunity by the papain-like protease domain of the severe acute respiratory syndrome coronavirus. J Biol Chem. 2007 Nov;282:32208–21. doi: 10.1074/jbc.M704870200 PMID: 17761676

32. Banerjee A, Zhang X, Yip A, Schulz KS, Irving AT, Bowdish D, et al. Positive selection of a serine residue in bat IRF3 confers enhanced antiviral protection. iScience. 2020 Mar;23(7):100958. doi: 10.1016/j.isci.2020.100958 PMID: 32179480

33. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Feb;579:270–3. doi: 10.1038/s41586-020-2012-7 PMID: 32015507

34. Needle D, Lountos GT, Waugh DS. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. Acta Cryst. 2015 Feb;D71:1102–11. doi: 10.1107/S1399004715003521 PMID: 25945576

35. Xue X, Yu H, Yang H, Xue F, Wu Z, Shen W, et al. Structures of two coronavirus main proteases: Implications for substrate binding and antiviral drug design. J Virol. 2008 Mar;82(5):2515–27. doi: 10.1128/JVI.02114-07 PMID: 18094151

36. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. EMBO J. 2002 Jul;21(13):3213–24. doi: 10.1093/emboj/cdf327 PMID: 12093723

37. Zhu X, Wang D, Zhou J, Pan T, Chen J, Yang Y, et al. Porcine deltacoronavirus nsp5 antagonizes type I interferon signaling by cleaving STAT2. J Virol. 2017 May;91(10):e00003–17. doi: 10.1128/JVI.00003-17 PMID: 28250121

38. Wang, D, Fang L, Shi Y, Zhang H, Gao L, Peng G, et al. Porcine epidemic diarrhea virus 3C-like protease regulates its interferon antagonism by cleaving NEMO. J Virol. 2016 Feb;90(4):2090–101. doi: 10.1128/JVI.02514-15 PMID: 26656704

39. Moustaqil M, Ollivier E, Chiu HP, Van Tol S, Rufolffi-Soto P, Stevens C, et al. SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. Emerg Microbes Infect. 2021 Dec;10(1):178–95. doi: 10.1080/22221751.2020.1870414

40. de Vries AAF, Horzinek MC, Rottier PJM, de Groot RJ. The genome organization of the *Nidovirales*: Similarities and differences between arteri-, toro-, and coronaviruses. Sem Virol. 1997 Feb;8(1):33–47. doi: 10.1006/smvy.1997.0104 PMID: 32288441

41. Ye S, Xia H, Dong C, Cheng Z, Xia X, Zhang J, et al. Identification and characterization of Iflavirus 3C-like protease processing activities. Virol. 2012 Jul;428(2):136–45. doi: 10.1016/j.virol.2012.04.002 PMID: 22534091

42. Kuyumcu-Martinez M, Belliot G, Sosnovtsev SV, Chang KO, Green KY, Lloyd RE. Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. J Virol. 2014 Aug;78(15):8172–82. doi: 10.1128/JVI.78.15.8172-8182.2004 PMID: 15254188

43. Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ. Substrate requirements of human rhinovirus 3C protease for peptide cleavage *in vitro*. J Biol Chem. 1990 Jun;265(16):9062–5. PMID: 2160953

44. Pallai PV, Burkhardt F, Skoog M, Schreiner K, Bax P, Cohen KA, et al. Cleavage of synthetic peptides by purified poliovirus 3C proteinase. J Biol Chem. 1989 Jun;264(17):9738–41. PMID: 2542331

45. Dougherty WG, Semler BL. Expression of virus-encoded proteinases: Functional and structural similarities with cellular enzymes. Microbiol Rev. 1993 Dec;57(4):781–822. PMID: 8302216

46. Tesar M, Marquardt O. Foot-and-mouth disease virus protease 3C inhibits cellular transcription and mediated cleavage of histone H3. Virology. 1990 Feb;174(2):364–74. doi: 10.1016/0042-6822(90)90090-e PMID: 2154880

47. Falk MM, Grigera PR, Gergmann IE, Zibert A, Multhaup G, Beck E. Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. J Virol. 1990 Feb;64(2):748–56. doi: 10.1128/JVI.64.2.748-756.1990 PMID: 2153239

48. Kliewer S, Dasgupta A. An RNA polymerase II transcription factor inactivated in poliovirus-infected cells copurifies with transcription factor TFIID. Mol Cell Biol. 1988 Aug;8(8):3175–82. doi: 10.1128/mcb.8.8.3175 PMID: 2850483

49. Clark ME, Dasgupta A. A transcriptionally active form of TFIIIC is modified in poliovirus-infected HeLa cells. Mol Cell Biol. 1990 Oct;10(10):5106–13. doi: 10.1128/mcb.10.10.5106 PMID: 2204807

50. Clark ME, Hammerle T, Wimmer E, Dasgupta A. Poliovirus proteinase 3C converts an active form of transcription factor IIIC to an inactive form: A mechanism for inhibition of host cell polymerase III transcription by poliovirus. EMBO J. 1991 Oct;10(10):2941–7. doi: 10.1002/j.1460-2075.1991.tb07844.x PMID: 1915271

51. Clark ME, Lieberman PM, Berk AJ, Dasgupta A. Direct cleavage of human TATA-binding protein by poliovirus protease 3C *in vivo* and *in vitro*. Mol Cell Bio. 1993 Feb;13(2):1232–7. doi: 10.1128/mcb.13.2.1232 PMID: 8380894

52. Shen Y, Igo M, Yalamanchili P, Berk AJ, Dasgupta A. DNA binding domain and subunit interactions of transcription factor IIIC revealed by dissection with poliovirus 3C protease. Mol Cell Biol. 1996 Aug;16(8):4163–71. doi: 10.1128/mcb.16.8.4163 PMID: 8754815

53. Joachims M, Etchison D. Poliovirus infection results in structural alteration of a microtubule-associated protein. J Virol. 1992 Oct;66(10):5797–804. doi: 10.1128/JVI.66.10.5797-5804.1992 PMID: 1326643

54. Joachims M, Harris KS, Etchison D. Poliovirus protease 3C mediates cleavage of microtubule-associated protein 4. Virology. 1995 Aug;211(2):451–61. doi: 10.1006/viro.1995.1427 PMID: 7645249

55. Thiel V, Ivanov KA, Putics A, Hertzig T, Schelle B, Bayer S, et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol. 2003 Sep;84(9):2305–15. doi: 10.1099/vir.0.19424-0 PMID: 12917450

56. Chuck CP, Chow HF, Wan DCC, Wong KB. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. PLOS One. 2011 Nov;6(11):e27228. doi: 10.1371/journal.pone.0027228 PMID: 22073294

57. Gao F, Ou HY, Chen LL, Zheng WX, Zhang CT. Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. FEBS Lett. 2003 Oct;553(3):451–6. doi: 10.1016/s0014-5793(03)01091-3 PMID: 14572668

58. Yang ZR. Mining SARS-CoV protease cleavage data using non-orthogonal decision trees: a novel method for decisive template selection. Bioinformatics. 2005 Jun;21(11):2644–50. doi: 10.1093/bioinformatics/bti404 PMID: 15797903

59. Kiemer L, Lund O, Brunak S, Blom N. Coronavirus 3CLpro proteinase cleavage sites: Possible relevance to SARS virus pathology. BMC Bioinform. 2004 Jun;5:72. doi: 10.1186/1471-2105-5-72 PMID: 15180906

60. Blom N, Hansen J, Blaas D, Brunak S. Cleavage site analysis in picornaviral polyproteins: Discovering cellular targets by neural networks. Protein Science. 1996 Sep;5(11):2203–16. doi: 10.1002/pro.5560051107 PMID: 8931139

61. Narayanan A, Wu X, Yang ZR. Mining viral protease data to extract cleavage knowledge. Bioinformatics. 2002 Jul;18(1):S5–13. doi: 10.1093/bioinformatics/18.suppl_1.s5 PMID: 12169525

62. Singh O, Su ECY. Prediction of HIV-1 protease cleavage site using a combination of sequence, structural, and physiochemical features. BMC Bioinform. 2016 Dec;17(Suppl 17):478. doi: 10.1186/s12859-016-1337-6 PMID: 28155640

63. Li X, Hu H, Shu L. Predicting human immunodeficiency virus protease cleavage sites in nonlinear projection space. Mol Cell Biochem. 2010 Jan;339(1–2):127–33. doi: 10.1007/s11010-009-0376-y PMID: 20054614

64. Rognvaldsson T, You L. Why neural networks should not be used for HIV-1 protease cleavage site prediction. Bioinformatics. 2004 Feb;20(11):1702–9. doi: 10.1093/bioinformatics/bth144 PMID: 14988129

65. You L, Garwicz D, Rognvaldsson T. Comprehensive bioinformatic analysis of the specificity of human immunodeficiency virus type 1 protease. J Virol. 2005 Oct;79(19):12477–86. doi: 10.1128/JVI.79.19.12477–12486.2005 PMID: 16160175

66. Cai YD, Chou KC. Artificial neural network model for predicting HIV protease cleavage sites in protein. Adv Eng Softw. 1998 Mar;29(2):119–28. doi: 10.1016/S0965-9978(98)00046-5

67. Manning T, Walsh P. The importance of physiochemical characteristics and nonlinear classifiers in determining HIV-1 protease specificity. Bioengineered. 2016 Mar–Apr;7(2):65–78. doi: 10.1080/21655979.2016.1149271 PMID: 27212259

68. Chown H. A comparison of machine learning algorithms for the prediction of hepatitis C NS3 protease cleavage sites. J Proteom Bioinform. 2019 Jul;12(5):89–95. doi: 10.35248/0974-276X.19.12.501

69. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019 Jan;47(D1):D505–15. doi: 10.1093/nar/gky1049 PMID: 30395287

70. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J. GenBank. Nucleic Acids Res. 2017 Jan;45(D1):D37–42. doi: 10.1093/nar/gkw1070 PMID: 27899564

71. Bukhari K, Mulley G, Gulyaeva AA, Zhao L, Shu G, Jiang J, et al. Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the *Coronavirinae*, the proposed genus Alphaletovirus. Virology. 2018 Nov;524:160–71. doi: 10.1016/j.virol.2018.08.010 PMID: 30199753

72. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Oct;7:539. doi: 10.1038/msb.2011.75 PMID: 21988835

73. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010 Jul;38(2):W695–9. doi: 10.1093/nar/gkq313 PMID: 20439314

74. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 2013 Jul;41(W1):W597–600. doi: 10.1093/nar/gkt376 PMID: 23671338

75. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. Genome Res. 2004 Jun;14:1188–90. doi: 10.1101/gr.849004 PMID: 15173120

76. Wu A, Wang Y, Zeng C, Huang X, Xu S, Su C, et al. Prediction and biochemical analysis of putative cleavage sites of the 3C-like protease of Middle East respiratory syndrome coronavirus. Virus Res. 2015 Oct;208:56–65. doi: 10.1016/j.virusres.2015.05.018 PMID: 26036787

77. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008 Nov;9:2579–605. doi: http://www.jmlr.org/papers/v9/vandermaaten08a.html

78. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011 Oct;12:2825–30. http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

79. Chou KC, Zhang CT, Kezdy FJ. A vector projection approach to predicting HIV protease cleavage sites in proteins. Proteins. 1993 Jun;16(2):195–204. doi: 10.1002/prot.340160206 PMID: 8332607

80. Bindewald E, Schneider TD, Shapiro BA. CorreLogo: an online server for 3D sequence logos of RNA and DNA alignments. Nucleic Acids Res. 2006 Jul;34:W405–11. doi: 10.1093/nar/gkl269 PMID: 16845037

81. Maes F, Vandermeulen D, Suetens P. Medical image registration using mutual information. Proc IEEE. 2003 Oct;91(10):1699–722. doi: 10.1109/JPROC.2003.817864

82. Kultys M, Nicholas L, Schwarz R, Goldman N, King J. Sequence Bundles: a novel method for visualizing, discovering and exploring sequence motifs. BMC Proc. 2014 Aug;8(Suppl 2):S8. doi: 10.1186/1753-6561-8-S2-S8 PMID: 25237395

83. Goetz DH, Choe Y, Hansell E, Chen YT, McDowell M, Jonsson CB, et al. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. Biochemistry. 2007 Jul;46(30):8744–52. doi: 10.1021/bi0621415 PMID: 17605471

84. Woo PCY, Huang Y, Lau SKP, Tsoi HW, Yuen KY. *In silico* analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. Microbiol Immunol. 2005 Oct;49(10):899–908. doi: 10.1111/j.1348-0421.2005.tb03681.x PMID: 16237267

85. Ma Y, Wu Y, Shaw N, Gao Y, Wang J, Sun Y, et al. Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. PNAS. 2015 Jul;112(30):9436–41. doi: 10.1073/pnas.1508686112 PMID: 26159422

86. Neuman BW, Chamberlain P, Bowden F, Joseph J. Atlas of coronavirus replicase structure. Virus Res. 2014 Dec;194:49–66. doi: 10.1016/j.virusres.2013.12.004 PMID: 24355834

87. Fang SG, Shen H, Wang J, Tay FPL, Liu DX. Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of *Coronavirus* infectious bronchitis virus is dispensable for viral replication in cultured cells. Virol. 2008 Sep;379(2):175–80. doi: 10.1016/j.virol.2008.06.038 PMID: 18678384

88. Santos MS, Soares JP, Abreu PH, Araujo H, Santos J. Cross-validation for imbalances datasets: Avoiding overoptimistic and overfitting approaches. IEEE Comp Intell Mag. 2018 Nov;13(4):59–76. doi: 10.1109/MCI.2018.2866730 PMID: 27295638

89. Fort S, Hu H, Lakshminarayanan B. Deep ensembles: A loss landscape perspective. arXiv [Preprint]. 2020 Jun;1912.02757

90. Kozlowski LP. Proteome-pI: proteome isoelectric point database. Nucleic Acids Res. 2017 Jan;45(D1):D1112–6. doi: 10.1093/nar/gkw978 PMID: 27789699

91. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009 Jan;4(1):44–57. doi: 10.1038/nprot.2008.211 PMID: 19131956

92. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009 Jan;37(1):1–13. doi: 10.1093/nar/gkn923 PMID: 19033363

93. Fan K, Wei P, Feng Q, Chen S, Huang C, Ma L, et al. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. J Biol Chem. 2004 Jan;279(3):1637–42. doi: 10.1074/jbc.M310875200 PMID: 14561748

94. Hegyi A, Ziebuhr J. Conservation of substrate specificities among coronavirus main proteases. J Gen Virol. 2002 Mar;83(3):595–99. doi: 10.1099/0022-1317-83-3-595 PMID: 11842254

95. Grum-Tokars V, Ratia K, Begaye A, Baker SC, Mesecar AD. Evaluating the 3C-like protease activity of SARS-coronavirus: Recommendations for standardized assays for drug discovery. Virus Res. 2008 Apr;133(1):63–73. doi: 10.1016/j.virusres.2007.02.015 PMID: 17397958

96. Fan K, Ma L, Han X, Liang H, Wei P, Liu Y, et al. The substrate specificity of SARS coronavirus 3C-like proteinase. Biochem Biophys Res Common. 2005 Apr;329(3):934–40. doi: 10.1016/j.bbrc.2005.02.061 PMID: 15752746

97. Ikemura T, Schwarze J, Makela M, Kanehiro A, Joetham A, Ohmori K, et al. Type 4 phosphodiesterase inhibitors attenuate respiratory syncytial virus-induced airway hyper-responsiveness and lung eosinophilia. J Pharmacol Exp Ther. 2000 Aug;294(2):701–6. doi: https://pubmed.ncbi.nlm.nih.gov/10900250/ PMID: 10900250

98. Mori I, Goshima F, Imai Y, Kohsaka S, Sugiyama T, Yoshida T, et al. Olfactory receptor neurons prevent dissemination of neurovirulent influenza A virus into the brain by undergoing virus-induced apoptosis. J Gen Virol. 2002 Sep;83(9):2109–16. doi: 10.1099/0022-1317-83-9-2109 PMID: 12185263

99. Thomander L, Aldskogius H, Vahlne A, Kristensson K, Thomas E. Invasion of cranial nerves and brain stem by herpes simplex virus inoculated into the mouse tongue. Ann Ontol Rhinol Laryngol. 1988 Sep;97(5):554–8. doi: 10.1177/000348948809700525 PMID: 2845851

100. Chilvers MA, McKean M, Rutman A, Myint BS, Silverman M, O'Callaghan C. The effects of coronavirus of human nasal ciliated respiratory epithelium. Eur Respir J. 2001 Dec;18(6):965–70. doi: 10.1183/09031936.01.00093001 PMID: 11829103

101. Kuek LE, Lee RJ. First contact: the role of respiratory cilia in host-pathogen interactions in the airways. Am J Physiol Lung Cell Mol Physiol. 2020 Oct;319(4):L603–19. doi: 10.1152/ajplung.00283.2020 PMID: 32783615

102. Li W, Li M, Ou G. COVID-19, cilia, and smell. FEBS J. 2020 Sep;287(17):3672–6. doi: 10.1111/febs.15491 PMID: 32692465

103. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020 Jul;583:459–68. doi: 10.1038/s41586-020-2286-9 PMID: 32353859

104. Lv X, Li Z, Guan J, Hu S, Zhang J, Lan Y, et al. Porcine hemagluttinating encephalomyelitis virus activation of the integrin α5β1-FAK-cofilin pathway causes cytoskeletal rearrangement to promote its invasion of N2a cells. J Virol. 2019 Mar;93(5):e01736–18. doi: 10.1128/JVI.01736-18 PMID: 30541856

105. Rudiger AT, Mayrhofer P, Ma-Lauer Y, Pohlentz G, Muthing J, von Brunn A, et al. Tubulins interact with porcine and human S proteins of the genus *Alphacoronavirus* and support successful assembly and release of infectious viral particles. Virol. 2016 Oct;497:185–97. doi: 10.1016/j.virol.2016.07.022 PMID: 27479465

106. Ohman T, Rintahaka J, Kalkkinen N, Matikainen S, Nyman TA. Actin and RIG-I/MAVS signaling components translocate to mitochondria upon influenza A virus infection of human primary macrophages. J Immunol. 2009 May;182(9):5682–92. doi: 10.4049/jimmunol.0803093 PMID: 19380815

107. Dohner K, Sodeik B. The role of the cytoskeleton during viral infection. Curr Top Microbiol. 2005 Feb;285:67–108. doi: 10.1007/3-540-26764-6_3 PMID: 15609501

108.   Naghavi MH, Walsh D. Microtubule regulation and function during virus infection. J Virol. 2017 Aug;91(16):e00538–17. doi: 10.1128/JVI.00538-17 PMID: 28615197

109.   Kristensson K, Lyche E, Roytta M, Svennerholm B, Vahlne A. Neuritic transport of herpes simplex virus in rat sensory neurons in vitro. Effects of substances interacting with microtubular function and axonal flow [nocodazole, taxol, and erythron-9-3-(2-hydroxynonyl)adenine]. J Gen Virol. 1986 Sep;67:2023–8. doi: 10.1099/0022-1317-67-9-2023 PMID: 2427647

110.   Solov'eva MF, Krispin TI, Shloma DV, Kalmakhelidze KA, Balandin IG. Effect of inhibitors that destroy cytoskeleton structures on the antiviral and antiproliferative activity of interferons. Vopr Virusol. 1988 May–Jun;33(3):309–14. PMID: 2459850

111.   Yi F, Guo J, Dabbagh D, Spear M, He S, Kehn-Hall K, et al. Discovery of novel small-molecule inhibitors of LIM domain kinase for inhibiting HIV-1. J Virol. 2017 Apr;91(13):e02418–16. doi: 10.1128/JVI.02418-16 PMID: 28381571

112.   Campbell EM, Nunez R, Hope TJ. Disruption of the actin cytoskeleton can complement the ability of Nef to enhance HIV-1 infectivity. J Virol. 2004 Jun;78(11):5745–55. doi: 10.1128/JVI.78.11.5745-5755.2004 PMID: 15140972

113.   Wolff G, Melia CE, Snijder EJ, Barcena M. Double-membrane vesicles as platforms for viral replication. Trends in Microbiol. 2020 Jun;1839. doi: 10.1016/j.tim.2020.05.009 PMID: 32536523

114.   Neuman BW, Angelini MM, Buchmeier MJ. Does form meet function in the coronavirus replicative organelle? Trends in Microbiol. 2014 Nov;22(11):642–7. doi: 10.1016/j.tim.2014.06.003 PMID: 25037114

115.   Miller E, Antonny B, Hamamoto S, Schekman R. Cargo selection into COPII vesicles is driven by the Sec24p subunit. EMBO J. 2002 Nov;21(22):6105–13. doi: 10.1093/emboj/cdf605 PMID: 12426382

116.   Mancias JD, Goldberg J. Structural basis of cargo membrane protein discrimination by the human COPII coat machinery. EMBO J. 2008 Oct;27(21):2918–28. doi: 10.1038/emboj.2008.208 PMID: 18843296

117.   Stagg SM, Gurkan C, Fowler DM, LaPointe P, Foss TR, Potter CS, et al. Structure of the Sec13/31 COPII coat cage. Nature. 2006 Jan;439:234–8. doi: 10.1038/nature04339 PMID: 16407955

118.   Nagesh PT, Husain M. Influenza A virus dysregulates host histone deacetylase 1 that inhibits viral infection in lung epithelial cells. J Virol. 2016 Apr;90(6):4614–25. doi: 10.1128/JVI.00126-16 PMID: 26912629

119.   Chen H, Qian Y, Chen X, Ruan Z, Ye Y, Chen H, et al. HDAC6 restricts influenza A virus by deacetylation of the RNA polymerase PA subunit. J Virol. 2019 Feb;93(4):e01896–18. doi: 10.1128/JVI.01896-18 PMID: 30518648

120.   Shulak L, Beljanski V, Chiang C, Dutta SM, Van Grevenynghe J, Belgnaou SM, et al. Histone deacetylase inhibitors potentiate vesicular stomatitis virus oncolysis in prostate cancer cells by modulating NF-κB-dependent autophagy. J Virol. 2014 Feb;88(5):2927–40. doi: 10.1128/JVI.03406-13 PMID: 24371063

121.   Feng Q, Su Z, Song S, Xu H, Zhang B, Yi L, et al. Histone deacetylase inhibitors suppress RSV infection and alleviate virus-induced airway inflammation. Int J Mol Med. 2016 Jul;38(3):812–22. doi: 10.3892/ijmm.2016.2691 PMID: 27460781

122.   Mosley AJ, Meekings KN, McCarthy C, Shepherd D, Cerundolo V, Mazitschek R, et al. Histone deacetylase inhibitors increase virus gene expression but decrease CG8+ cell antiviral function in HTLV-1 infection. Blood. 2006 Dec;108(12):3801–7. doi: 10.1182/blood-2006-03-013235 PMID: 16912225

123.   Kaminskyy V, Zhivotovsky B. To kill of be killed: how viruses interact with the cell death machinery. J Intern Med. 2010 May;267:473–82. doi: 10.1111/j.1365-2796.2010.02222.x PMID: 20433575

124.   Spencer CA, Kruhlak MJ, Jenkins HL, Sun X, Bazett-Jones DP. Mitotic transcription repression *in vivo* in the absence of nucleosomal chromatin condensation. J Cell Bio. 2000 Jul;150(1):13–26. doi: 10.1083/jcb.150.1.13 PMID: 10893252

125.   Banerjee S, An S, Zhou A, Silverman RH, Makino S. RNase L-independent specific 28S rRNA cleavage in murine coronavirus-infected cells. J Virol. 2000 Oct;74(19):8793–802. doi: 10.1128/jvi.74.19.8793-8802.2000 PMID: 10982321

126.   Slavov N, Semrau S, Airoldi E, Budnik B, van Oudenaarden A. Differential stoichiometry among core ribosomal proteins. Cell Rep. 2015 Nov;13(5):865–73. doi: 10.1016/j.celrep.2015.09.056 PMID: 26565899

127.   Plant EP, Rakauskaite R, Taylor DR, Dinman JD. Achieving a golden mean: Mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. J Virol. 2019 Apr;84(9):4330–40. doi: 10.1128/JVI.02480-09 PMID: 20164235

128.   Park H, Subramaniam AR. Inverted translational control of eukaryotic gene expression by ribosome collisions. PLOS Biol. 2019 Sep;17(9):e3000396. doi: 10.1371/journal.pbio.3000396 PMID: 31532761

129.   Joazeiro CAP. Mechanisms and functions of ribosome-associated protein quality control. Nat Rev Mol Cell Biol. 2019 Jun;20(6):368–83. doi: 10.1038/s41580-019-0118-2 PMID: 30940912

130.   Siegal V, Walter P. Elongation arrest is not a prerequisite for secretory protein translocation across the microsomal membrane. J Cell Biol. 1985 Jun;100(6):1913–21. doi: 10.1083/jcb.100.6.1913 PMID: 2581979

131.   Rottier P, Armstrong J, Meyer DI. Signal recognition particle-dependent insertion of coronavirus E1, an intracellular membrane glycoprotein. J Biol Chem. 1985 Aug;260(8):4648–52. PMID: 2985561

132.   Young JC, Andrews DW. The signal recognition particle receptor alpha subunit assembles co-translationally on the endoplasmic reticulum membrane during an mRNA-encoding translation pause *in vitro*. EMBO J. 1996 Jan;15(1):172–81. doi: 10.1002/j.1460-2075.1996.tb00345.x PMID: 25855820

133.   Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. Front Immunol. 2019 Sep;9(2039):1–16. doi: 10.3389/fimmu.2018.02039 PMID: 30250470

134.   Roy M, Viginier B, Saint-Michel E, Arnaud F, Ratinier M, Fablet M. Viral infection impacts transposable element transcript amounts in *Drosophila*. PNAS. 2020 Jun;117(22):12249–57. doi: 10.1073/pnas.2006106117 PMID: 32434916

135.   Wada M, Lokugamage KG, Nakagawa K, Narayanan K, Makino S. Interplay between coronavirus, a cytoplasmic RNA virus, and nonsense-mediated mRNA decay pathway. PNAS. 2018 Oct;115(43):e10157–66. doi: 10.1073/pnas.1811675115 PMID: 30297408

136.   Wang W, Xiog L, Wang P, Wang F, Ma Q. Major vault protein plays important roles in viral infection. IUBMB Life. 2020 Apr;74(4):624–31. doi: 10.1002/iub.2200 PMID: 31769934

137.   Steiner E, Holzmann K, Pirker C, Elbling L, Micksche M, Sutterluty F, et al. The major vault protein is responsive to and interferes with interferon-γ-mediated STAT1 signals. J Cell Sci. 2006 Oct;119:459–69. doi: 10.1242/jcs.02773 PMID: 16418217

138.   Li F, Chen Y, Zhang Z, Ouyang J, Wang Y, Yan R, et al. Robust expression of vault RNAs induced by influenza A virus plays a critical role in suppression of PKR-mediated innate immunity. Nucleic Acids Res. 2015 Dec;43(21):10321–37. doi: 10.1093/nar/gkv1078 PMID: 26490959

139.   Bellon M, Nicot C. Regulation of telomerase and telomeres: Human tumor viruses take control. J Natl Cancer Inst. 2008 Jan;100(2):98–108. doi: 10.1093/jnci/djm269 PMID: 18182620

140.   Reghunathan R, Jayapal M, Hsu LY, Chng HH, Tai D, Leung BP, et al. Expression profile of immune response genes in patients with severe acute respiratory syndrome. BMC Immunol. 2005 Jan;6:2. doi: 10.1186/1471-2172-6-2 PMID: 15655079

141.   Berlutti F, Pantanella F, Natalizi T, Frioni A, Paesano R, Polimeni A, et al. Antiviral properties of lactoferrin—A natural immunity molecule. Molecules. 2011 Aug;16(8):6992–7018. doi: 10.3390/molecules16086992 PMID: 21847071

142.   Adusumilli NC, Zhang D, Friedman JM, Friedman AJ. Harnessing nitric oxide for preventing, limiting and treating the severe pulmonary consequences of COVID-19. Nitric Oxide. 2020 Oct;103:4–8. doi: 10.1016/j.niox.2020.07.003 PMID: 32681986

143.   Perrone LA, Belser JA, Wadford DA, Katz JM, Tumpey TM. Inducible nitric oxide contributes to viral pathogenesis following highly pathogenic influenza virus infection in mice. J Infect Dis. 2013 May;207(10):1576–84. doi: 10.1093/infdis/jit062 PMID: 23420903

144.   Wan G, Zhaorigetu S, Liu Z, Kaini R, Jiang Z, Hu CAA. Apolipoprotein L1, a novel Bcl-2 homology domain 3-only lipid-binding protein, induces autophagic cell death. J Biol Chem. 2008 Aug;282(31):21540–9. doi: 10.1074/jbc.M800214200 PMID: 18505729

145.   Peng Y, Wan L, Fan C, Zhang P, Wang X, Sun J, et al. Cholesterol metabolism—Impact for SARS-CoV-2 infection prognosis, entry, and antiviral therapies. medRxiv [Preprint]. 2020 Apr. doi: 10.1101/2020.04.16.20068528

146.   Rebello CJ, Kirwan JP, Greenway FL. Obesity, the most common comorbidity in SARS-CoV-2: is leptin the link? Int J Obes (Lond). 2020 Jul;44:1810–7. doi: 10.1038/s41366-020-0640-5 PMID: 32647360

147.   Zhang AJX, To KKW, Li C, Lau CCY, Poon VKM, Chan CCS, et al. Leptin mediates the pathogenesis of severe 2009 pandemic influenza A (H1N1) infection associated with cytokine dysregulation in mice with diet-induced obesity. J Infect Dis. 2013 Apr;207(8):1270–80. doi: 10.1093/infdis/jit031 PMID: 23325916

148.   Chesarino NM, McMichael TM, Yount JS. E3 ubiquitin ligase NEDD4 promotes influenza virus infection by decreasing levels of the antiviral protein IFITM3. PLOS Pathog. 2015 Aug;11(8):e1005095. doi: 10.1371/journal.ppat.1005095 PMID: 26263374

149.   Shi G, Ozog S, Torbett BE, Compton AA. mTOR inhibitors lower an intrinsic barrier to virus infection mediated by IFITM3. PNAS. 2018 Oct;115(43):e10069–78. doi: 10.1073/pnas.1811892115 PMID: 30301809

150.   Xu Q, Zhu N, Chen S, Zhao P, Ren H, Zhu S, et al. E3 ubiquitin ligase Nedd4 promotes Japanese encephalitis virus replication by suppressing autophagy in human neuroblastoma cells. Sci Rep. 2017 Mar;7:45375. doi: 10.1038/srep45375 PMID: 28349961

151.   Yu J, Li M, Wilkins J, Ding S, Swartz TH, Esposito AM, et al. IFITM proteins restrict HIV-1 infection by antagonizing the envelope glycoprotein. Cell Rep. 2015 Oct;13:145–56. doi: 10.1016/j.celrep.2015.08.055 PMID: 26387945

152.   Zhu X, He Z, Yuan J, Wen W, Huang X, Hu Y, et al. IFITM3-containing exosome as a novel mediator for anti-viral response in dengue virus infection. Cell Microbiol. 17(1):105–18. doi: 10.1111/cmi.12339 PMID: 25131332

153.   Huang IC, Bailey CC, Weyer JL, Radoshitzky SR, Becker MM, Chiang JJ, et al. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. PLOS Pathog. 2011 Jan;7(1):e1001258. doi: 10.1371/journal.ppat.1001258 PMID: 21253575

154.   Zhao X, Sehgal M, Hou Z, Cheng J, Shu S, Wu S, et al. Identification of residues controlling restriction versus enhancing activities of IFITM proteins on entry of human coronaviruses. J Virol. 2018 Mar;92(6):e01535–17. doi: 10.1128/JVI.01535-17 PMID: 29263263

155.   Hachim MY, Al Heialy S, Hachim IY, Halwani R, Senok AC, Maghazachi AA, et al. Interferon-induced transmembrane protein (IFITM3) is upregulated explicitly in SARS-CoV-2 infected lung epithelial cells. Front Immunol. 2020 Jun;11:1372. doi: 10.3389/fimmu.2020.01372 PMID: 32595654

156.   Yang B, Kumar S. Nedd4 and Nedd4-2: closely related ubiquitin-protein ligases with distinct physiological functions. Cell Death Differ. 2009 Jun;17:68–77. doi: 10.1038/cdd.2009.84 PMID: 19557014

157.   Han Z, Lu J, Liu Y, Davis B, Lee MS, Olson MA, et al. Small-molecule probes targeting the viral PPxY-host Nedd4 interface block egress of a broad range of RNA viruses. J Virol. 2014 Apr;88(13):7294–306. doi: 10.1128/JVI.00591-14 PMID: 24741084

158.   An H, Krist DT, Statsyuk AV. Crosstalk between kinases and Nedd4 family ubiquitin ligases. Mol BioSyst. 2014 Jan;10:1643–57. doi: 10.1039/C3MB70572B PMID: 24457516

159.   Maaroufi H. SARS-CoV-2 encodes a PPxY late domain motif that is known to enhance budding and spread in enveloped RNA viruses. bioRxiv [Preprint]. 2020 Apr. doi: 10.1101/2020.04.20.052217

160.   Isaacson MK, Ploegh HL. Ubiquitination, ubiquitin-like modifiers, and deubiquitination in viral infection. Cell Host Microbe. 2009 Jun;5:559–70. doi: 10.1016/j.chom.2009.05.012 PMID: 19527883

161.   Zinngrebe J, Montinaro A, Peltzer N, Walczak H. Ubiquitin in the immune system. EMBO Rep. 2013 Nov;15(1):28–45. doi: 10.1002/embr.201338025 PMID: 24375678

162.   Steimer L, Klostermeier D. RNA helicases in infection and disease. RNA Biol. 2012 Jun;9(6):751–771. doi: 10.4161/rna.20090 PMID: 22699555

163.   Sharma A, Boris-Lawrie K. Determination of host RNA helicases activity in viral replication. Methods Enzymol. 2012 Jun;511:405–35. doi: 10.1016/B978-0-12-396546-2.00019-X PMID: 22713331

164.   Umate P, Tuteja N, Tuteja R. Genome-wide comprehensive analysis of human helicases. Commun Integr Biol. 2011 Jan–Feb;4(1):118–37. doi: 10.4161/cib.4.1.13844 PMID: 21509200

165.   Xu L, Khadijah S, Fang S, Wang L, Tay FPL, Liu DX. The cellular RNA helicase DDX1 interacts with coronavirus nonstructural protein 14 and enhances viral replication. J Virol. 2010 Sep;84(17):8571–83. doi: 10.1128/JVI.00392-10 PMID: 20573827

166.   Chem JY, Chen WN, Poon KMV, Zheng BJ, Lin X, Wang YX, et al. Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. Arch Virol. 2009 Feb;154(3):507–12. doi: 10.1007/s00705-009-0323-y PMID: 19224332

167.   Zhang Z, Kim T, Bao M, Facchinetti V, Jung SY, Ghaffari AA, et al. DDD1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. Immunity. 2011 Jun;34(6):866–78. doi: 10.1016/j.immuni.2011.03.027 PMID: 21703541

168.   Meier-Stephenson V, Mrozowich T, Pham M, Patel TR. DEAD-box helicases: the Yin and Yang roles in viral infections. Biotechnol Genet Eng Rev. 2018 Apr;34(1):3–32. doi: 10.1080/02648725.2018.1467146 PMID: 29742983

169.   Biembengut IV, de Souza TACB. Coagulation modifiers targeting SARS-CoV-2 main protease Mpro for COVID-19 treatment: an *in silico* approach. Mem Inst Oswaldo Cruz. 2020 Jun;115:e200179. doi: 10.1590/0074-02760200179 PMID: 32490889

170.   Spiezia L, Boscolo A, Poletto F, Cerruti L, Tiberio I, Campello E, et al. COVID-19-related severe hypercoagulability in patients admitted to intensive care unit for acute respiratory failure. Thromb Haemost. 2020 Jun;120(6);998–1000. doi: 10.1055/s-0040-1710018 PMID: 32316063

171.   Ji HL, Zhao R, Matalon S, Matthay MA. Elevated plasmin(ogen) as a common risk factor for COVID-19 susceptibility. Physiol Rev. 2020 Jul;100(3):1065–75. doi: 10.1152/physrev.00013.2020 PMID: 32216698

172.   Sottrup-Jensen L, Sand O, Kristensen L, Fey GH. The α-macroglobulin bait region: Sequence diversity and localization of cleavage sites for proteinases in five mammalian α-macroglobulins. J Biol Chem. 1989 Sep;264(27):15781–9. PMID: 2476433

173.   Gettins PGW, Hahn KH, Crews BC. α2-macroglobulin bait region variants: A role for the bait region in tetramer formation. J Biol Chem. 1995 Jun;270(23):14160–7. doi: 10.1074/jbc.270.23.14160 PMID: 7539801

174.   Meyer M, Jaspers I. Respiratory protease/antiprotease balance determined susceptibility to viral infection and can be modified by nutritional antioxidants. Am J Physiol Lung Cell Mol Physiol. 2015 Apr;308:L1189–201. doi: 10.1152/ajplung.00028.2015 PMID: 25888573

175.   De Bont CM, Boelens WC, Pruijn GCM. NETosis, complement, and coagulation: a triangular relationship. Cell Mol Immunol. 2019 Jan;16(1):19–27. doi: 10.1038/s41423-018-0024-0 PMID: 29572545

176.   Noris M, Benigni A, Remuzzi G. The case of complement activation in COVID-19 multiorgan impact. Kidney Intl. 2020 Aug;98(2):314–22. doi: 10.1016/j.kint.2020.05.013 PMID: 32461141

177.   Agrawal P, Nawadkar R, Ojha H, Kumar J, Sahu A. Complement evasion strategies of viruses: An overview. Front Microbiol. 2017 Jun;8:1117. doi: 10.3389/fmicb.2017.01117 PMID: 28670306

178.   Eddie Ip WK, Chan KH, Law HKW, Tso GHW, Kong EKP, Wong WHS, et al. Mannose binding lectin in severe acute respiratory syndrome coronavirus infection. J Infect Dis. 2005 May;191(10):1697–704. doi: 10.1086/429631 PMID: 15838797

179.   Narasuraju T, Tang BM, Herrmann M, Muller S, Chow VTK, Radic M. Neutrophilia and NETopathy as key pathologic drivers of progressive impairment in patients with COVID-19. Front Pharmacol. 2020 Jun;11:870. doi: 10.3389/fphar.2020.00870 PMID: 32581816

180.   Zou Y, Yalavarthi S, Shi H, Gockman K, Zuo M, Madison JA, et al. Neutrophil extracellular traps in COVID-19. JCI Insight. 2020 Apr;5(11):e138999. doi: 10.1172/jci.insight.138999 PMID: 32329756

181.   Stahel PF, Barnum SR. Complement inhibition in coronavirus disease (COVID)-19: A neglected therapeutic option. Front Immunol. 2020 Jul;11:1661. doi: 10.3389/fimmu.2020.01661 PMID: 32733489

182.   De Costa MG, Poppelaars F, van Kooten C, Mollnes TE, Tedesco F, Wurzner R, et al. Age and sex-associated changes of complement activity and complement levels in healthy Caucasian population. Front Immunol. 2018 Nov;9:2664. doi: 10.3389/fimmu.2018.02664 PMID: 30515158

183.   Khomich OA, Kochetkov SN, Bartosch B, Ivanov AV. Redox biology of respiratory viral infections. Viruses. 2018 Aug;10(8):392. doi: 10.3390/v10080392 PMID: 30049972

184.   Schneider D, Ganesan S, Comstock AT, Meldrum CA, Mahidhara R, Goldsmith AM, et al. Increased cytokine response of rhinovirus-infected airway epithelial cells in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2010 Aug;182(3):332–40. doi: 10.1164/rccm.200911-1673OC PMID: 20395558

Graphical abstract
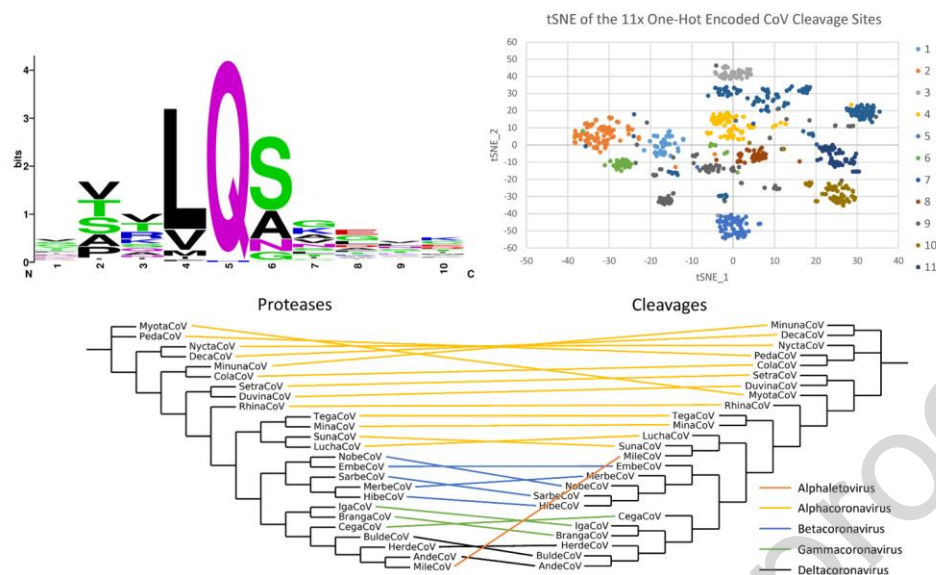
Prescott CRediT Authorship Contribution Statement

Lucas Prescott: Conceptualization, Methodology, Data Curation, Software, Validation, Formal Analysis, Writing, Visualization.

Prescott Conflicts of Interest Statement

The author certifies that there is NO affiliation with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript titled "SARS-CoV-2 3CLpro whole human proteome cleavage analysis prediction and enrichment/depletion analysis." This project was conducted in spare time outside my full-time employment and without any funding.

Prescott Highlights

- Gathered a large 3CLpro cleavage dataset.
- Optimized multiple machine learning (ML) models to predict 3CLpro cleavages.
- Applied the best model to the whole human proteome and performed enrichment/depletion analysis.
- Discussed the possible consequences of predicted cleavages in many tissues and cellular pathways and proposed more than a dozen potential therapeutic targets.